Deriving and Applying Facet Views of the Dewey Decimal Classification Scheme to Enhance Subject Searching in Library OPACs

Amanda Jayne Tinker
BA (Hons.) MSc. MCLIP

A thesis submitted to the University of Huddersfield in partial fulfilment of the requirements for the degree of Doctor of Philosophy

April 2005

ABSTRACT

Classification is a fundamental tool in the organisation of any library collection for effective information retrieval. Several classifications exist, yet the pioneering Dewey Decimal Classification (DDC) still constitutes the most widely used scheme and international *de facto* standard. Although once used for the dual purpose of physical organisation and subject retrieval in the printed library catalogue, library classification is now relegated to a singular role of shelf location. Numerous studies have highlighted the problem of subject access in library online public access catalogues (OPACs). The library OPAC has changed relatively little since its inception, designed to find what is already known, not discover and explore.

This research aims to enhance OPAC subject searching by deriving facets of the DDC and populating these with a library collection for display at a View-based searching OPAC interface. A novel method is devised that enables the automatic deconstruction of complex DDC notations into their component facets. Identifying facets based upon embedded notational components reveals alternative, multidimensional subject arrangements of a library collection and resolves the problem of disciplinary scatter. The extent to which the derived facets enhance users' subject searching perceptions and activities at the OPAC interface is evaluated in a small-scale usability study.

The results demonstrate the successful derivation of four fundamental facets (Reference Type, Person Type, Time and Geographic Place). Such facet derivation and deconstruction of Dewey notations is recognised as a complex process, owing to the lack of a uniform notation, notational re-use and the need for distinct facet indicators to delineate facet boundaries. The results of the preliminary usability study indicate that users are receptive to facet-based searching and that the View-based searching system performs equally as well as a current form fill-in interface and, in some cases, provides enhanced benefits. It is concluded that further exploration of facet-based searching is clearly warranted and suggestions for future research are made.

ACKNOWLEDGEMENTS

I would firstly like to thank the late Dr. Steve Pollitt, my Director of Studies for the first part of this project. He inspired me to undertake this PhD research and provided tremendous support; very sadly, he did not live to see the completion of this thesis. I also wish to extent my grateful thanks to Dr. Phil Marsden, my supervisor and, subsequently, Director of Studies for the second half of this research; his experience and pragmatism has been invaluable.

I would like to thank staff at OCLC, particularly Joan Mitchell, Diane Vizine-Goetz and Julianne Beall, for their interest and permission to utilise the Dewey Decimal Classification Scheme in this research project. I am also indebted to my colleagues at the University of Huddersfield Library and Computing Centre. Particular thanks to Sue White for kindly agreeing to sponsor the usability study and allowing me valuable time to 'write up', Lynn Stevens for covering my reduced working hours and Dave Pattern for his assistance with the transaction log capture. A big thank you also to Patrick Braekevelt, Pat Woods and Howard Lockwood at View-based Systems Limited and everyone who participated in the usability study.

Finally, a special thanks to my family and friends for their interest, support and endless encouragement.

Dedicated to the Memory of Dr. Arthur Steven Pollitt

1950 - 2004

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

Ever since humans first began to record and collect knowledge, there has been a need to devise techniques for organising and providing access to these growing information stores. Long before the advent of the computer, classification emerged as a viable means for providing such access by physically arranging library items in a 'helpful order' by subject matter, assigning a notation to group like with like. The growth and diversification of library collections led to their documentation and retrieval by the additional tool of the library catalogue. In early printed catalogues, particularly in Great Britain, classification played a prominent role in subject retrieval, in the form of the 'classified catalogue'. However, as bibliographic standards were sought and batch-mode, intermediary driven computer technology emerged, the rival alphabetic 'dictionary catalogue' predominated and became the model for the subsequent Online Public Access Catalogue (OPAC).

Today's library OPACs are therefore borne from an exact-match, known-item search model, with little focus on addressing the problems of subject searching, highlighted over the years by numerous user studies (Bates, 1977, 1986, 1989; Borgman, 1986, 1996; Hancock, 1987; Markey, 1989; Kuhlthau, 1999). Such studies revealed that users have problems in matching their own vocabulary to that of the system, narrowing or broadening searches and formulating often vague information needs into a specific search statement. The library OPAC has changed relatively little since its inception, designed to find what is already known, not discover and explore. Despite the display of classificatory structures being suggested as a means of assisting subject searching and its recent discovery and re-invention by Internet and Semantic Web communities for this purpose, classification still lies dormant within commercial library OPACs, relegated from subject search tool to shelf-locating device. This change in the role of library classification has perhaps led to an underestimation of its potential. Today, libraries primarily rely upon Library of Congress Subject Headings for subject retrieval in OPACs. However, classificatory notations encode a similar

richness of information, coupled with additional browsing capabilities of an underlying knowledge structure, and yet still remain untapped. The value of such structures for searching and browsing may depend on the actual scheme employed, since there are several available library classification schemes, each with their own advantages and disadvantages. However, the most famous and widely used scheme is the Dewey Decimal Classification (DDC).

Created by Melvil Dewey in 1876 and now in its 22nd edition, the Dewey Decimal Classification Scheme is used in over 200,000 libraries in 135 countries (OCLC, 2004) and is incorporated into the bibliographic records of significant national and research libraries such as the Library of Congress and the British Library, whose records are adopted throughout the world. Although retaining its predominant 'mark and park' role of physical shelf organisation, the knowledge structures of Dewey have been previously applied, to a limited degree, at the system interface for subject retrieval. Applications include an experimental OPAC (Markey, 1989) and more recently its use in Internet subject gateways. However, such applications have unnecessarily replicated the restrictions of physical linear shelf arrangements, which is at odds with the multidimensional nature of many of today's information and library items and the diverse needs of users. This is a missed opportunity.

If classifications such as Dewey are to be applied in the electronic domain of a library OPAC for enhancing subject searching, then they should not be governed by the same constraints as a linear physical collection. Physical applications of a classification are restricted by the need for a single relative location for a particular library item; a location typically determined by the first component in the classificatory notation, namely subject discipline. In a growing universe of knowledge, the subject content of library items can embody several different components, even crossing disciplinary boundaries. Although not explicitly represented in the library shelf order, such embedded notational components are available for deriving alternative arrangements at the electronic interface. Provision of alternative arrangements by subject dimensions (facets) such as 'time' and 'place' would enable a history student, for example, to view and collate a library collection by subordinated subject facets such as time period, which would otherwise be notationally and physically 'scattered' across the classification scheme and library building. By exploiting the potential of

Dewey's embedded subject facets, we may move current classificatory subject access beyond a single 'best fit' shelf arrangement, designed solely for physical access, to one which affords many 'tailored', multidimensional, representations at the OPAC interface. However, the provision of such alternative classificatory arrangements and multidimensional subject access is currently beyond the scope of today's form fill-in OPAC interfaces, which rely upon specification and discourage browsing. This potential is, however, provided for by an approach to information retrieval called View-based searching (Pollitt, 1996; 1997). View-based searching is based upon the principles of faceted classification, utilising knowledge structures at the system interface with implicit Boolean searching. If the rich notational structure of Dewey and its embedded facets were exploited within a View-based Searching interface, then we may go some way towards achieving the improved subject searching, browsability and multidimensional access to library collections that is so lacking in our current library OPACs.

This research aims to:

- Enhance OPAC subject searching by devising a method for deriving facets of the Dewey Decimal Classification Scheme and populating these with a library collection for display at a View-based Searching interface.
- Examine the extent to which the derived facets enhance users' subject searching perceptions and activities.

This current doctoral research is part of a larger project to apply the conceptual structures of the Dewey Decimal Classification Scheme at the interface of a View-based Searching OPAC of the University of Huddersfield library collection. The above components were therefore 'givens' in terms of this PhD research. Thus, the DDC and the faceted approach of View-based Searching is the inherited research platform. However, although the View-based Searching software requires faceted knowledge structures at its interface in order to realise its full potential, the DDC is not conceived according to this faceted form, as will be demonstrated in the forthcoming chapters. Manipulation of the Dewey Decimal Classification Scheme is therefore clearly needed and this became the primary focus for the doctoral research, coupled with a view to addressing the previously mentioned, well-documented,

problems of subject access in library OPACs. Providing more in-depth study within the overarching project, this research therefore seeks to investigate and exploit the faceted potential of the DDC by deconstructing its classificatory notations, moving classification from a predominately linear to multidimensional tool at the online interface. This 're-fashioning' of Dewey into a more faceted form will fully exploit the potential of View-based Searching for addressing the problems of subject access, providing a browsable OPAC interface, more conducive to the exploratory nature of subject searching than current specification-led approaches. Moreover, once implemented, the devised method for deriving and populating Dewey facets could be applied to any other library collection that has been classified by the DDC.

The thesis is organised into nine further chapters. The first two chapters provide contextual information on the two main components of this research, namely library classification and its changing role in the library catalogue, arguing that an historical legacy has contributed to the current research problem of the relegation of classification and concurrent subject searching difficulties in library OPACs. Whilst libraries are forgoing their oldest knowledge organisation tool for information retrieval in library OPACs, conversely, its value is being recognised in wider information contexts. In addition to examining the demise of classification in library catalogues, chapter 3 demonstrates how it is being utilised for organising and browsing digital resources on the Web and its relationship to the various classification-based tools that are creating a 'meaningful' Semantic Web. The remainder of the thesis then returns to and seeks to address this relegation of classification by libraries and the related problems of subject access and a need for more browsable OPAC interfaces. Chapter 4 examines the problem in more detail, expounding a rationale for deconstructing and deriving additional Dewey facets for exploitation at the OPAC interface. Chapter 5 then presents the procedural method devised for achieving this, followed by the resultant four case studies of Chapters 6 and 7, which explicate, document and analyse this method in practice. The results of this facet derivation are then applied and demonstrated within a View-based Searching OPAC interface in Chapter 8, and the deconstruction process intrinsically evaluated. Further evaluation, but from a user perspective, is reported in Chapter 9, which presents the method and results of a small-scale usability study. Finally, the thesis is concluded and recommendations are made for future research.

CHAPTER 2

CLASSIFICATION AND THE ORGANISATION OF LIBRARIES

> It may safely be assumed that everything in the universe is a member of some class, but on the first appearance the universe appears so great and complex that it is chaos - a tangle of things to which man had no clue unless he provides himself with some sort of map. This map of things is a convenient expression for a classification scheme, for we cannot reason, even in the simplest manner unless we can identify and relate - that is classify - things.
>
> Arthur Maltby (1975, p.16)

2.1 Introduction

Classification is ubiquitous. We use it constantly to make sense of the world, to impose order on otherwise disarray. Since time began, all living things have relied on an ability to classify ('friend or foe', 'edible or inedible') for survival. As humans, we developed an unique capacity for expressing such distinctions through language, providing us with conceptual representations for grouping, separating and perceiving relationships in the world and its population by endless dimensions such as age, sex, occupation, nationality, colour of hair, colour of eyes and so on and on. We attach such 'labels' to make the world seem smaller, more manageable. Just as we developed such groupings to organise our view of the world, libraries, as documentary repositories of this externalised, recorded and shared universal knowledge, were in need of a similar systematic arrangement. The basic action of classification, physically grouping and separating documents according to such 'labels', provided a powerful organisational tool.

The purpose of this chapter is to demonstrate how classification can be regarded as the earliest tool for the organisation and retrieval of information, and to place modern day bibliographic classification into context. The Dewey Decimal Classification scheme (DDC) is introduced and situated within the context of alternative universal classification schemes and the reasons why this nineteenth century scheme continues to predominate are explored. With the increasing textual complexity of the twentieth and interdisiciplinary challenges of the twenty-first century, the more recent and more flexible faceted schemes appeared aptly placed to meet such a challenge, yet their

adoption was relatively minimal. Pioneering enumerative schemes, attempting to predict and explicitly list all conceptual relationships within the universe of knowledge, continue to prevail, with Dewey now the *de facto* standard. Although this chapter focuses on classification in the physical realm of shelf organisation (now, although not always, its primary function, as will be revealed in the following Chapter 3), underpinning this chapter and entire thesis is a view that we can increase the flexibility of such, often cumbersome, physical schemes by freeing them from such constraints: by manipulation in the online domain.

The chapter firstly defines and situates classification as a means of grouping and arranging physical documents on library shelves. This traditional role of shelf organisation is then explored in the context of a brief history of classificatory theory and practice from its ancient roots through to the modern day. Five classificatory 'eras' are defined and the various schemes within these are discussed. It is argued that despite twentieth century developments of classification, which today more closely correspond to knowledge production and meet the needs of users, schemes with their roots in the nineteenth century still and will continue to predominate. As such, new electronic approaches should be explored to offset their limitations and realise their full potential.

2.2 Classification: Definitions and Purpose

Classification, in its broadest sense, is generally accepted as a process of grouping like with like according to a particular characteristic (Buchanan, 1979; Marcella and Newton, 1994; Hunter, 2000). Its use pervades our everyday lives; as young children we classify to understand the world, developing a personalised mental map, which expands and is refined as we encounter new concepts throughout life; at school we are streamed into classes of like ability; and in our adult lives, our places of work are organised into departments, which conduct similar activities. Classification is all-pervasive.

This current research, however, pertains to the grouping of documents, a classificatory activity that holds its own idiosyncrasies and complexities, as Mills (1960, p.1) points out:

> A major task of the librarian is to see that maximum use is made of the collection of which he is in charge - "Every reader his book, every book its reader". This collection consists of books and other records; unlike packets of sugar in a grocer's shop or machine components in a factory's store, each book is not only unique but exhibits relations of considerable complexity to other books. This complexity is hardly surprising for the printed record has been for centuries, and still is, a fundamental mechanism by which human thought is communicated, and the thought content is in some respects co-extensive with that of human knowledge.

In this the opening of his classic text 'A modern outline of library classification', here Mills conveys the magnitude of the task in hand. Unlike the broad groupings we see in everyday life, documents may embody multiple concepts; for example, a book entitled *Scandinavia - ceramics and glass in the 20th century* contains concepts of place ('Scandinavia'), materials ('ceramics', 'glass') and time ('20th century'), which potentially could be related to a series of other documents exhibiting each, some or all of these characteristics. The challenge of organising, retrieving and relating documents representing this multidimensional universe of knowledge stimulated the development of highly specific bibliographic classification schemes.

Bibliographic classification provides a means for the systematic arrangement of documents according to the degree of likeness by subject matter. It indicates relationships by collocating documents with certain similar characteristics whilst, concurrently, separating them from those that are different; what Olson (2001) terms 'the duality of sameness and difference'. Imposing such structure and organisation upon extensive and diverse library collections ensures their optimum use, enabling a user to retrieve documents on a specific topic more efficiently and effectively, without which there would be 'chaos', as Maltby (1975, p.17) envisages:

> A huge room or building full of books and pamphlets, in fact, having regard to the variety of subjects represented in them and the many forms and sizes they take, about the nearest representation to chaos that we can imagine. That is to say, if they are not classified in some way apart from their mere separation from things which are not books. Unless they are well classified, we cannot discover without immense loss of time what books there are on chemistry, history, theology, fluid mechanics, or transport, for example.

Traditionally, and still predominantly, the primary purpose of bibliographic classification is to provide a physical organisation of documents on library shelves, providing a 'helpful order' whereby similar materials are grouped together (Foskett, 1996). However, as will be discussed in subsequent chapters, it is also used for

organisation and retrieval in classified catalogues and bibliographies, and more recently as a browsing tool in online databases and Internet gateways.

This section has presented generic definitions and briefly considered the typical roles of classification. In order to explore this further, when examining the theoretical underpinnings and development of classification, several writers (e.g. Mills, 1997; McIlwaine and Williamson, 1999) have stressed the value of adopting and learning from an historical approach. In his introductory address at the Sixth International Study Conference on Classification Research (to mark the fortieth anniversary of the seminal Dorking Conference) Jack Mills clearly encapsulates this:

> Such an approach, like classification itself, has as one of its central functions the establishment of a wider context in which events should be viewed if our view is not to be distorted by too narrow a focus.
>
> Mills (1997, p.1)

The following sections review classification in greater depth and in light of an historical perspective.

2.3 Early Classification

It is perhaps not surprising that when humans first began to record and collect knowledge to pass onto future generations, they drew upon, externalised and developed their natural predisposition for classification as a means for organising and providing access to these growing information stores. McArthur (1986, p.32) calls this human need to categorise and order 'the taxonomic urge':

> Classification and thematisation were already inherent in society at large. Then as now, society and its view of the world were replete with order, system and stratification. Communities have their ranks, castes and guilds, armies their divisions, priesthoods their hierarchies, merchants their inventories, farmers their fields and boundaries, architects their various structural conventions, and the gods their pantheons.

Although in today's terms comparatively simplistic, these ancient library classifications constituted the first attempts at providing a helpful physical arrangement by grouping together whilst simultaneously disassociating information-bearing entities on the basis of likeness.

Fascinating early evidence of subject organization was apparent in the ruins of the Assyrian library of Assurbanipal. Johnson and Harris (1976) describe how some

thirty thousand clay tablets were physically separated and housed in different 'subject rooms'; one relating to history and government; another geography; and others for laws and legal decisions, commercial records, mythology and legends, and the sciences and pseudo sciences. These clay tablets were housed in earthen jars and ordered on shelves; interestingly, the authors find remnants of the first classificatory notation in the form of a tag that designated the jar, shelf, and subject room in which the tablet was located. However, the accuracy of this description is cast into doubt when others claim evidence for only two main classes: the earth and the heavens with subdivision within these (Maltby, 1975; Marcella and Newton, 1994), and when Miksa (1994, p.144) acknowledges that texts were organised but asserts "there were no divisions of knowledge in the modern sense of the word".

There is consensus, however, regarding the earliest recorded classification scheme, which was created by Callimachus for the library of the Pharoahs at Alexandria between 260-240 BC. Thompson (1977) notes that there is some disagreement in the literature as to the exact nature of the main classes and their subdivisions but it is clear that they were based on the affiliation of the writer, thus creating an implicit rather than explicit subject classification. Richardson's version of these classes seems to be the most widely accepted (cited in Maltby, 1975, p.110):

Poets
Law-makers
Philosophers
Historians
Rhetoricians
Miscellaneous writers

A more explicit subject orientation than that of the Alexandria library can be seen in the classifications of ancient China. The Seven Epitomes of Liu Hsin (1st century AD) took its inspiration from philosophy there, which was beginning to view knowledge holistically; following a general summary section, the scheme grouped documents into six classes: arts, poetry, philosophy, the military, science, technology and medicine, all with further subdivisions (Miksa, 1994). This scheme was replaced during the Wei and Tsin dynasties (3rd - 5th centuries) by a scheme named the *Cheng Mo* with its system of four main classes (classics, philosophy, history and literature), which remarkably survived until the twentieth century. A similarly holistic and long

serving subject scheme was developed in ancient Islamic libraries between the 8th and 11th centuries.

In the western world, the fall of the Roman Empire resulted in widespread cultural decline, considerably reducing the size and scope of library collections. Libraries were typically confined to monasteries, having just a few hundred manuscripts often contained in book chests. Consequently, in the Middle Ages there was no real impetus for organisation by classification, and even the larger collections might only be given a simple binary division into "theological works and secular ones; between Latin works and those in other languages; or between textbooks and serious tomes" (Johnson and Harris 1976, p.108).

When collections did begin to flourish with the establishment of university libraries, the growing number of secular works were organised to reflect the educational curriculum of the day: the 'seven liberal arts', each designated relative importance under the groupings of the lower order *trivium* (grammar, rhetoric and logic) and the higher *quadrivium* (arithmetic, geometry, astronomy and music). To this day, the Bodleian library, founded in 1602, still bears testament to this legacy, evidenced by Latin inscriptions to this effect over the doors in the Old Schools Quadrangle.

2.4 Philosophical, Scientific and Pragmatic Classifications

Studies of the historical development of bibliographic classification (Miksa, 1994, 1998; Wellisch and Smiralia, 1993; Thompson, 1977; Sayers, 1975; Dolby, 1979; Marcella and Newton, 1994) have noted parallel classificatory movements between the 15th and 18th centuries; philosophical classifications of knowledge and the sciences, and the encyclopaedic movement; alongside more pragmatic classifications of library and bookseller collections. Researchers have traditionally argued that philosophical/scientific classifications were a major impetus for and influence upon the development of modern library classification (Miksa, 1994); however, although still subject to debate, the connection is now thought not to be quite so strong (Miksa, 1998).

The emergence and proliferation of philosophical classifications conceptualised the existence of a 'universe of knowledge', embracing a recognition and drive to demonstrate a holistic view, the inter-relatedness of subjects and their systemisation within a hierarchical (genus-species) structure. This interest in classifying knowledge and the sciences arose from a surge in knowledge production and artistic creativity during the 15th and 16th century Renaissance period, straining the former pedagogic *trivium* and *quadrivium* organisation and leading several individuals to devise new, more sophisticated schemes; the most influential being that of Francis Bacon in his treatise the 'Advancement of Learning' (1605) (Miksa, 1994; Sayers, 1975; Dolby 1979). McArthur (1986, p.110) believes Bacon's ambitious rationale was "to re-organise all of human knowledge", resulting in a hierarchical structure that conceptualised the entire universe of knowledge and divided then subdivided until complete coverage was achieved. Bacon's classification schematised knowledge by three major categories *Memory, Imagination* and *Reason*, as follows:

Memory	**History**	**Natural history** **Civil history** **Ecclesiastical history, Theology**
Imagination	**Poetry**	**Literature** **The Arts**
Reason	**Philosophy**	**The Sciences** **Civil philosophy, Sociology, Politics** **Economic Science**

Figure 2.1: Bacon's Classification of Knowledge (cited in Marcella and Newton, 1994, p.67)

However, Bacon's classification was purely philosophical not bibliographical, as Maltby (1975. p.116) summarises:

> The *Advancement of Learning* is therefore the history of the record of thought as it was at the date of its appearance: but it is also rather a discussion of the state of knowledge than of the books or methods themselves. Bacon rarely, and only then by allusive methods, mentions the names of books.

Nevertheless, it has been characterised as 'profoundly influential' (Maltby, 1975, p.117; Marcella and Newton, 1994, p.67). Bacon's classification, despite not being conceived for bibliographic purposes, was purportedly the basis for the early

classification of the Bodleian library, the first classification of the Library of Congress and "became the basis of countless library catalogs (*sic*) from the 17th to the 19th centuries" (Miksa, 1994, p.146). Moreover, it is said to have influenced Melvil Dewey in construction of the first version of the nineteenth century Dewey Decimal Classification Scheme, the Main Class order of which is said to be based on the inverted Baconian classification of William Torey Harris via Edward William Johnston (Comaromi, 1983).

The 18th century saw the rise of the European encyclopaedia movement, which pursued a view of knowledge as a unified, connected whole. Their ethos was didactic:

> They envisioned good public education as the activity of developing, through proper intellectual exercise, the mental faculties which each person has at birth. Such exercise, facilitated through reading the ideas of the best thinkers in the light of the place of their ideas in a systematically arranged universe of knowledge, would produce mentally cultivated people capable of contributing to sound government and societal advance.
>
> Miksa (1994, p.146)

The first encyclopaedia, 'Cyclopedia, or Universal Dictionary of the Arts and Sciences', was created by Ephraim Chambers in 1728; later followed by Diderot and D'Alembert's 'Encyclopédie, ou dictionnaure Rasisonné des Sciences, des Arts et des Méteiers, par une Société de Gens de Lettres', the construction of which, the authors attribute, was strongly influenced by Francis Bacon (Maltby, 1975, p.117). The intention of such encyclopaedic works was to provide a thematic approach to knowledge; however, today this has been largely superseded by an alphabetic arrangement, often the two approaches are combined yet still with alphabetic precedence. For example, the Propaedia outline of knowledge in Encyclopaedia Britannica can be used in conjunction to the primary alphabetic Macro and Micro-paedia volumes. Similarly, Roget's Thesaurus was conceived according to a classified approach, influenced by Bacon and the biological taxonomies of the day but, as MacArthur (1986, p.153) notes, was later replaced by an alphabetic arrangement:

> Peter Mark Roget brought out his *Thesaurus*...to serve as a converse to the dictionary and to display associations among words in a format that approximated to the mind. It is both ironic, therefore, and indicative of the sheer gravitational pull of ABC ordering that a number of books calling themselves Roget's Thesaurus now exist in standard alphabetical arrangement.

As shall be seen in Chapter 3, the classified library catalogue was to suffer a similar fate.

Following the classificatory efforts of the philosophers and encyclopaedists, the scientific community was also bitten by the 'taxonomic urge'. The late eighteenth century and early nineteenth century saw the creation of classifications in botany and zoology. Dolby (1979) claims that during this time there was an increased interest in an inductive philosophy of science. With science in its embryonic stages and lacking established status, such inductive philosophy was preoccupied with producing elaborate classifications to characterise and justify the field of study. Consequently, scientific classifications proliferated in the nineteenth century to "...enormous proportions ...when anyone who was anybody wrote a treatise on the topic" (Miksa, 1998, p.34).

The classifications discussed in this section thus far have all been abstract, philosophical in nature, in a sense, aimed at perfecting intrinsic 'ideal' classifications rather than ones motivated by an extrinsic need such as the organisation of physical 'things'. Although the ancient classifications discussed in section were developed to fulfil such an extrinsic purpose, their simplicity cannot merit the status of a library classification scheme in the modern sense of the word. Developments in France during the 15th to early 19th century, however, were to give more recent classification this practical outlet.

In 1498, the Frenchman Aldus Manutius devised what can be considered the first modern classification scheme for library arrangement (Marcella and Newton, 1994; Maltby, 1975). Such a scheme is therefore distinct in its pragmatic purpose when compared to the theoretical classifications of the philosophers and scientists, a distinction frequently made in the literature by theorists such as Edward Edwards (Maltby, 1975) and Wyndham Hulme and his notion of 'literary warrant'[1]. Furthermore, the invention and propagation of printing in the 15th and 16th centuries created a need for booksellers to maintain orderly records, leading the Paris booksellers, under the direction of Ismael Bouilleau and later Jacques-Charles Brunet,

to develop a practical classification scheme; a scheme which was to influence the organisation of French bibliographies and libraries, notably, and still today, the Bibliothèque Nationale (Wellisch and Smiraglia, 1993; Maltby, 1975).

Two distinct approaches towards classification therefore developed between the 15th and early 19th century; on the one hand, the philosophical classifications of knowledge and the sciences; and on the other, the practical classifications of libraries, booksellers' catalogues and bibliographies. Miksa (1998, p. 45) summarises whilst questioning the accepted view of a strong connection between the two:

> Library classificationists took some ideas from the movement to classify knowledge and the sciences ... But unlike the proponents of the movement to classify knowledge and the sciences, who felt they had to be apologists for the role and status of science in the modern world, library classificationists simply assumed that the conclusions of the knowledge classificationists were both good and even necessary. But apart from that, they had their own task of getting knowledge to library patrons in the form of books. And that task had everything to do with practicality and very little to do with philosophical speculation about knowledge.

At the beginning of the 20th century, the drive to produce philosophical classifications of knowledge and the sciences, which had so characterised the 19th century, died out (Dolby, 1975). Science had achieved its societal status, no longer feeling a need for justification by mass enumerations of the entire universe of knowledge. Miksa (1998) notes that from then on such activities became the sole preserve of librarians, which, consequently, becomes the focus of the remainder of this chapter.

2.5 Nineteenth century classification: pragmatism, hierarchical structures, and enumeration

The modern bibliographic classification schemes of today's libraries emerged in the late nineteenth century. For the first time, libraries allowed users to roam freely amongst their collections, motivating a need for explicit and helpful arrangements, which did not rely on the intimate or idiosyncratic knowledge of the librarian. Prior to this, the purpose of arrangement had been efficient known-item retrieval by a librarian, with closed access collections typically being organised by non-subject

---

[1] Hulme (cited in Mills, 1960) argued that classification schemes should not be based on philosophical principles but be designed according to the subject content of pre-existing literature.

characteristics, an arrangement not particularly conducive to user browsing. In fact, this is a legacy still evident in older libraries today; for example the National Art Library at the Victoria and Albert Museum, London, organises its closed access collection according to size and date of acquisition, with items retrieved by library assistants upon user request, and its open-access reference collection by the subject-based Dewey Decimal Classification Scheme. Therefore, analogous with the schemes of the fifteenth century Paris booksellers, the origin of modern library classification was instigated and driven by a pragmatic need.

Early modern library classification was characterised by hierarchical structures which, arguably influenced by the foregoing philosophical classifications previously discussed, aimed at a complete and extensive documentation of the entire universe of knowledge. Such classifications adopt a deductive approach, viewing knowledge as a unified whole, then dividing and subdividing it into increasingly specific categories (or classes) until complete coverage is achieved. In so doing, various hierarchical relationships are created whereby "the subordinate class is wholly contained by the superordinate class" (Broughton, 2001a) such as the aforementioned taxonomic/genus-species (e.g. Domestic Cats/Felines - Siamese Cats, Manx Cats etc.) and also partonomic/whole-part relationships (e.g. Human Body - head, arms, legs etc.). These nineteenth and early twentieth hierarchical classification schemes attempted to list explicitly all such relationships; hence, this drive for such inclusive documentation led them to be known as enumerative schemes. It should, however, be noted that, although a hierarchical structure and enumeration do tend to co-occur, they are not synonymous, since the former describes the type of relationship and the latter the extent to which these relationships are listed in the classification.

As a result of this nineteenth century drive for extensive documentation, coupled with the rigidity of hierarchical structures, such classifications can be unwieldy in enumeration and be prone to inefficient repetition of common concepts. For example, if we return to the earlier book title *'Scandinavia - ceramics and glass in the 20th century'* a purely hierarchical/enumerative classification might provide a ready-made construction for such an item by fully documenting recurring concepts such as Scandinavia (geographic place) and 20th century (time period) within the primary subject focus/discipline, perhaps here 'Art and Design'. Similarly, such recurrent

concepts and even structures (e.g. all the possibly relevant geographic areas or time periods) would be supplied *again* at all the other possibly relevant points in the classification. Prediction and possibility therefore underlie purely hierarchical/enumerative classification, since relationships need to be anticipated, pre-coordinated and documented in advance. Although the literary warrant of published output is an available source for determining relationships, it looks to the past, one step behind future required relationships that need to be incorporated into an existing hierarchical structure. In recent years, traditionally hierarchical/enumerative classifications have followed classificatory trends and techniques to minimise such repetition and increase flexibility, as will be discussed.

Three of the five major universal classification schemes used in libraries today were conceived during the late nineteenth and early twentieth century, namely the pioneering Dewey Decimal Classification (DDC) and focus of this current research, the related Universal Decimal Classification (UDC) and the highly enumerative Library of Congress Classification (LCC). The following sections outline the two schemes that are pertinent to the current research: the DDC and the UDC.

2.5.1 The Dewey Decimal Classification Scheme

In 1873, a twenty-one year old student library assistant at Amherst College, Massachusetts, recognised the need to make collections accessible to the end-user by devising what was to become the most widely used classification scheme in existence: the Dewey Decimal Classification. Melvil Dewey, regarded as one of the founding fathers of modern librarianship (Metcalf, 1983), first published his classification '*A Classification and Subject index for Cataloging and Arranging the Books and Pamphlets of a Library'* anonymously in 1876. In fact, Dewey's name did not appear in the title of the classification until the thirteenth edition in 1932, published a year after his death (Miksa, 1998). Decimal arrangement for library collections predates Dewey but previously books were limited to a 'fixed location' on subject numbered shelves, restricting a book's movement and making new additions cumbersome. Melvil Dewey pioneered the practice of a 'single relative location' by way of labelling each book with a decimal notation, allowing them to be moved freely around the library building yet be easily returnable to their ordered sequence. Dewey therefore

revolutionised the arrangement of library collections by recognising that the location of a book did not necessarily equate to a fixed 'physical space' on the library shelves but, more flexibly and hospitably, an "intellectual space in a classification system" (Chan et al, 1996, p.2), an intellectual space capable of indefinite expansion for accommodating new subjects, courtesy of the decimal notation. Now in its 22nd edition, published July 2003, it is used in over 135 countries (OCLC, 2004) and has become a *de facto* international standard, since 1980 included in Library of Congress MARC records, cataloguing in publication data and various national bibliographies (Foskett, 1982, p.339). Although, perhaps traditionally associated with public libraries, Dewey has expanded its remit to university and even national libraries, recently adopted by the British Library for organising its open-access collections.

The DDC began life as an enumerative classification with notations for all (including composite) subjects being explicitly listed in the main body of the classification scheme, known as the *Schedules* (Chan et al, 1996, p.36). These Schedules are based on subject discipline, consisting of ten main classes:

000	Generalities
100	Philosophy and psychology
200	Religion
300	Social sciences
400	Language
500	Natural sciences and mathematics
600	Technology (Applied Sciences)
700	The arts Fine and decorative arts
800	Literature and rhetoric
900	Geography and history

As mentioned in section 2.3, Dewey was influenced by Bacon, via W T Harris, in his choice of main classes (Comaromi, 1983). The choice and order of classificatory main classes received much attention in the nineteenth century, when there was a drive to emulate the order of nature prescribed by the scientific classifications and taxonomies. Consequently, classification schemes have been judged as to whether or not they present a logical order of main classes. The DDC has often been subject to criticism in this respect, namely in its separation of the related subject disciplines such as language (400) and literature (800) (Foskett, 1996), which necessitates a break in the numerical sequence if such subjects are to be presented within close proximity on the library shelves.

These Main Classes are then divided into a second hierarchical level known as 'Divisions', for example:

200	Religion
210	Philosophy and theory of religion
220	Bible
230	Christianity
240	Christian moral and devotional theology
250	Christian orders and local church
260	Social and ecclesiastical theology
270	History of Christianity and Christian Church
280	Christian denominations and sects
290	Comparative religion and other religions

Followed by a third, termed 'Sections', for example:

270	History of Christianity and Christian Church
271	Religious orders in church history
272	Persecutions in church history
273	Doctrinal controversies and heresies
274	History of Christianity in Europe
275	History of Christianity in Asia
276	History of Christianity in Africa
277	History of Christianity in North America
278	History of Christianity in South America
279	History of Christianity in other areas

These top three hierarchical levels are known as the 'Summaries'. The above examples of the 'Summaries' demonstrate successive subordination and increasing specificity both structurally and notationally. The fact that there is generally a direct correspondence between Dewey's hierarchical structure and its similarly hierarchical notation, means that the notation is expressive and not just a simple ordinal device; however, as is discussed below, the value of notational expressivity within a classification scheme is subject to debate. The first incarnation of the DDC comprised just the above three levels (Mitchell, 2001), but, in subsequent editions, the growth of knowledge led to the inclusion of further subdivision. The choice, here, of the 200 Religion class exemplifies a common criticism of the DDC, namely its Anglo-American bias. As can be seen from the second-level 'Division', the focus is almost exclusively upon Christianity, with only the *290 Comparative religion and other religions* explicitly devoted to non-Christian religions. Although there are options within Dewey available to libraries with predominantly non-Christian collections, Oh and Yeo (2001) note that these cater only for a single alternative religion, not collections such as those in Korea in which religious diversity predominates. The

existence of such bias epitomises the scheme's nineteenth century origins and, accordingly, predominately westernised worldview. Thus, in response to such criticism, initiatives to reduce this Christian bias have been incorporated into the 22nd edition of the DDC. For example, most significantly, 'Christianity' has been relocated from seemingly representing the entire religion class at the Main Class level 200 to Division status at 230.

Another criticism is the rigidity of ten divisions, prescribed by the decimal notation. As the above third-level summary shows, when considering the history of Christianity in particular geographic areas, the remaining two continents are contained within a 'catch-all' division of *'279 History of Christianity in other areas'* and not each given a dedicated caption and notation. However, this might be influenced by literary warrant, in that there may be less published output, relating to Christianity and the geographic areas within this division; accordingly, each country does not merit its own section. In addition to this 'bundling' into a catchall section, the rigidity of ten divisions, on occasions, can create an anomalous structure. Therefore, despite, in most cases, achieving notational expressivisity, when faced with more than ten potential subdivisions, the DDC hierarchy can break down. Pollitt (1998) cites the DDC geographic area hierarchy as one example of how notation can distort logical structure:

--4 Europe Western Europe
--41 British Isles
--42 England and Wales
--43 Central Europe Germany
--44 France and Monaco
--45 Italian Peninsula and adjacent islands Italy
--46 Iberian Peninsula and adjacent islands Spain
--47 Eastern Europe Russia
--48 Scandinavia
--49 Other parts of Europe

As the above shows, the British Isles and England and Wales are presented as equivalencies (coordinates), at the same hierarchical level, when, in fact, there should be a partitive hierarchical relationship between the two, with England and Wales being subdivisions of the British Isles. Furthermore, again we see the ubiquitous ninth 'other' catchall subdivision. This sequential, as opposed to hierarchical, structure is even more apparent at lower levels:

--41 British Isles
--411 Scotland
--412 Northeastern Scotland
--413 Southeastern Scotland
--414 Southwestern Scotland
--415 Ireland
--416 Ulster
--417 Republic of Ireland (Eire)
--418 Leinster
--419 Munster

Again, it can be seen that even within the British Isles, of which in this representation England and Wales are confusingly absent, parts of Scotland are notationally given the same status as their 'parent' subdivision Scotland. Similarly, Ulster and Eire are equivalencies of overarching Ireland, and Leinster and Munster of Eire. In a personal communication with Ross Trotter (2000), then Chair of the Library Association (now Chartered Institute of Library and Information Professionals) Dewey Committee and responsible for several draft revisions of parts of the DDC, he attributes such preference for sequential, rather than hierarchical structure, to substantial literary warrant and a desire for notational brevity:

> There is a conflict between keeping the notational hierarchy regular and keeping the numbers as short as possible. A prime example of this is the treatment of Panthera (Leo) in the new Life Sciences schedule at 599.755. Strictly, Leopard, Snow Leopard, Lion Tiger and Cheetah should all be subdivisions of 599.755 and so have the notation 599.755x. But because of the huge literary warrant for the latter three they have instead been given coordinate notation at 599.75x to keep the numbers as short as possible, while the use of the notation to accurately display hierarchy goes out of the window.

Similarly, in relation to the geographic area example above, Trotter (2000) offers the following 'solution':

- 41 Scotland and Ireland
- 411 Scotland
- 411x Divisions of Scotland
- 415 Ireland
- 415x Divisions of Ireland

Furthermore, he explains how this would preclude the use of notations –412 to -414 and –416 to –419, which would be seen as a "waste of scare notation" and result in eight, as opposed to the current four, digit notations for major cities with substantial literary warrant such as Glasgow and Edinburgh. He concludes that, due to the inherent limit of ten subdivisions within Dewey's decimal notation, in certain overcrowded Schedule areas, it is inevitable that some subordinate concepts will be allotted coordinate notations.

This raises the question as to whether classifications should be attempting to achieve notational expressivity at all. As Mills (1960), Maltby (1975) and Foskett (1996), amongst others, stress, quite correctly, that the primary purpose of a notation is to create a linear order for shelf arrangement; however:

> Notation should reflect order, not determine it...The systematic sequence of topics is the essence of library classification. Notation is only the mechanism which maintains that sequence; it should be considered only after the problems of sequence have been decided.
> Mills (1960, p.38)

In light of the above discussion, we could perhaps then perceive Dewey, with is restricted decimal base for subdivision, as a case of, so to speak, the 'tail wagging the dog'. For example, in her examination of the pros and cons of an expressive notation, Broughton (1999) notes that computer science is squeezed into the low-level class mark 001.64; however, the 20th edition of Dewey promoted this subject to the third level sections 004 – 006 (Miksa, 1998). Thus, it still does not merit the rank of Main Class, despite now having equivalent subject discipline status in the majority of universities. Instead, it remains a subdivision of the inexplicitly captioned 'Generalities' Main Class, perhaps due to the limit of ten Main Classes in the DDC, historical legacy and the potential disruption caused by such a dramatic re-classification. From the outset, Broughton (1999) claims that the need for an expressive notation is 'arguable', believing that users are generally unaware that a notation such as the DDC encodes a hierarchical structure and relationships. This is probably the case, when the user only views the notation in a sequential order at the library shelves but if the notation encoded hierarchical structures that could be utilised as an online browsing tool, then the quality of notational expressivity would be invaluable. However, the caveat being that any anomalous structures or disparities between conceptual structure and notation, as demonstrated in the examples above, would also become more apparent. This tension between classification for shelf organisation, maintaining a linear order, and classification for online information retrieval, enabling broadening and narrowing of searches, is explored further in the subsequent Chapter 3. However, even before the advent of the computer, there was awareness amongst the burgeoning scientific community of certain shortcomings in Dewey and the related challenges of an information retrieval context. In 1895, this

prompted two Belgian bibliographers, Paul Otlet and Henri La Fontaine, to propose an extension of the DDC, then known as the 'Brussels Expansion' and latterly, and more popularly, as the Universal Decimal Classification Scheme.

2.5.2 The Universal Decimal Classification Scheme

Despite not originating as a tool for shelf-arrangement, the Universal Decimal Classification (UDC), described as "the world's foremost multilingual classification scheme for all fields of knowledge, a sophisticated indexing and retrieval tool" (UDC Consortium, 2002), is now used as such in thousands of libraries worldwide. First published in France in 1905, Otlet and Fontaine gained permission from Melvil Dewey to re-develop the DDC into a more precise classification to meet the demands of a growing, and increasingly specialised, body of scientific literature (Miksa, 1998). This published output was no longer just in the form of books, but also periodicals, reports, standards, conference papers and even patents. The UDC, however, was not developed as a means for arranging such documents on library shelves but for information retrieval in a classified catalogue/bibliography known as the Répertoire Bibliographique Universel. Miksa (1998) argues that the complexities and practical act of classifying such specialised materials, in contrast to the general library context of Melvil Dewey, led Otlet and Fontaine to re-conceptualise the notion of a subject and contribute to the development of classificatory theory.

Otlet and Fontaine maintained the accepted nineteenth century hierarchical view of knowledge as a unified whole but they were dealing with more complex subjects, containing several components rather than a single identifiable topic, which were also buried deep within the classificatory hierarchy (Miksa, 1998). Faced with this specialised context and the desire to represent complex subjects within the classification and its notation, Otlet and Fontaine supplemented Dewey's numerical notation with various symbols. This changed the DDC from a pure to a mixed notation, creating structural (clearly demonstrating a relationship between two or more conceptual components) as well as hierarchical expressivity (Broughton, 1999). These supplementary symbols enabled the specificity needed for information retrieval, as opposed to shelf arrangement, to be incorporated into the UDC notation, as follows (McIlwaine, 1993, pp.36-44):

	Linking Devices
+	Links two or more non-consecutive UDC notations, when an enumerated notation does not exist.
/	Links the first and last notations of a series to indicate a broad concept or a range of consecutive concepts.
:	Links two or more notations, the order of which can be reversed, depending on the required emphasis.
::	Links a superordinate (preceding the double colon) to a subordinate concept.
[ ]	Used as a sub grouping device when a number of the above symbols are utilised.
	Recurring Concepts
=	Language of the item, followed by a notation to specify which.
(0…)	Form of presentation, e.g. dictionary etc.
(1-9)	Place. Items about a particular geographic area, specified in brackets.
(=…)	Ethnic Grouping and Nationality
"…"	Time

Figure 2.2: Notational symbols in the Universal Decimal Classification Scheme (UDC)

Figure 2.2 illustrates the impressive structural expressivity capabilities of the UDC notation, created by its various symbols for representing compound subjects and unambiguously representing recurrent concepts. Thus, in contrast to the DDC with its sole 0 digit, conceptual boundaries are clearly identifiable and delineated. For example, if we consider a possible UDC notation and an actual DDC notation from the Library of Congress Online Catalogue for a book entitled *Mining, metallurgy, and minting in the Middle Ages*:

UDC [622 + 669 + 737]"04/14"
Mining (622), Metallurgy (669), Numismatics (737), Middle Ages ("04/14") i.e.400-1499/5th to 15th century

DDC 662.0902
Mining (622), historical, geographic, persons treatment (09), 6th to 15th centuries (02)

The above example illustrates equivalent classificatory notations for the same book title. As can be seen from this, the UDC notation incorporates all of the four subject components, namely 'mining', 'metallurgy', 'numismatics' (minting) and the 'middle ages', all of which are clearly demarcated by various notational symbols. The UDC time period component is particularly informative, enclosed in quotation marks that unequivocally identify it as such and the digits themselves are literally mnemonic, in that they explicitly convey the actual time period, 400-1499. In contrast, the DDC example is much less inclusive and informative. Firstly, it represents just two components, the 622 subject component, 'mining', and the 0902 time period component, '6th-15th century'. Thus, unlike the UDC, the DDC cannot cater for such a diversely complex subject; therefore, a choice has to be made as to which subject discipline will be represented. In the DDC, this is governed by various factors, primarily author emphasis yet also instructions within the Schedules in the form of preference notes or tables (Chan et al, 1996). Secondly, the DDC notation is comparatively less structurally expressive. It only has the zero digit for separating components, with no indication as to their relationship, and the identification and meaning of recurring components, such as the 0902 time period component in the above example, is less apparent.

The capabilities of the UDC, shown in the above examples, for analysing, combining and notationally representing the separate components of a complex subject moved the act of classification forward from being a predominately enumerative task to one that also incorporated the flexibility of synthesis: an analytico synthetic classification. This synthesis, or building, of classificatory compounds is accomplished by re-using and combining Schedule notations or by appending re-current notations from a set of 'Auxiliary Tables', as in the case of the 'Middle Ages' time period component. Today, UDC has nine such tables, which enable re-current notations to be listed just once, rather than enumerated at all the predicted locations within the Schedules, creating a more economical and flexible scheme. This is not to say that Dewey was completely unaware of the economies of synthesis when developing his own pioneering scheme. As will be discussed at the end of Section 2.6 and also in Chapter 4, Dewey was aware of re-current concepts as early as the first edition of the DDC (Miksa, 1998) and, today, the scheme has six auxiliary tables for synthesising notations. Despite this, relative to the UDC, Dewey's notational devices for

accomplishing and representing such synthesis are comparatively limited. However, this does not mean that every notation in the UDC is built (synthesised) by the classifier; primary combinations are still pre-coordinated and enumerated but, in contrast to the DDC, each component is a recognisable and meaningful constituent and thus the notation can be more readily taken apart, reflecting the classification's motivation as an information retrieval tool. However, although the representation of such detail and complexity is useful in the information retrieval domain, long UDC notations can be regarded as being rather unwieldy and overly detailed for purposes of linear shelf arrangement in a general library. Therefore, as will become increasingly apparent throughout this thesis, there is a constant tension between the needs of classification of shelf arrangement and classification for information retrieval.

Through this recognition and notational representation of complex subjects, the UDC advanced classificatory theory, although not intentionally, as Miksa (1999) notes, but, like Dewey, as a by-product of devising solutions to a pragmatic problem. In Dewey's case, this was to provide a helpful arrangement of books on shelves; whereas, in the case of Otlet and Fontaine, it was organisation and retrieval within a printed bibliography. The analytico-synthetic devises of the UDC pointed the way forward for their twentieth century successors, where the impetus was primarily governed by theoretical development not pragmatism.

2.6 Twentieth century classification: theoretical development, facet analysis and synthesis

The twentieth century brought significant advances in classification. The growth and specialisation of knowledge was making mass enumerations difficult, stimulating an increased need to represent complex subjects, efficiently and effectively. The 19th century view of knowledge as a static entity, which could be taken in its entirety and subdivided until all conceptual relationships had been predicted and documented in advance, was soon to change.

In 1924, an Indian mathematician named Shiyali Ramamrita Ranganathan was appointed librarian at the University of Madras (Foskett, 1996, p.315). Having no previous experience of librarianship, he was required to study library science at

University College London. Here, whilst studying classification, he became increasingly dissatisfied by the rigidity of the 'top-down' approaches adopted by the DDC and even the UDC, akin to Maltby (1975, p.39):

> It can be said, indeed, that an enumerative classification, with its strict attempt at a 'genus to species' type of progression, is very like a photograph, or portrait, of a person. As a photograph shows an individual at a particular time and from a particular angle, so the enumerative classification arrests knowledge at a given stage of its development and organizes it rigidly in a selected form

Maintaining the photography analogy, it could be said that Ranganathan envisaged the flexibility of a digital not an analog photograph, where its subject *could* be shown from different angles. He began formulating his own ideas, based on a view of knowledge as a dynamic and multidimensional entity, which should not be divided and subdivided from the top downward but should begin at the bottom, with the individual entity; an entity which could be examined from various perspectives or facets (Broughton, 2001b).

Ranganthan first published his Colon Classification Scheme in 1933, with six subsequent editions in quick succession between 1939-1963, and a seventh edition in 1987 (Satija, 1997; Singh, 1999). Compared to the earlier established universal classification schemes, today, the Colon Classification is not widely used. Broughton (2001b) notes that the last known British application, at Christ's College Cambridge, was being replaced by the heavily enumerative Library of Congress Classification, and the Colon Classification is even lacking support from its roots in the Indian Sub-Continent (Singh, 1999). Possible reasons as to why will be explored later. However, despite this lack of application of the Colon Classification, the methodology of facet analysis, on which it is based, was to have a profound and lasting influence on modern bibliographic classification.

Facet analysis begins by analysing a subject area (what may constitute a Main Class) for representative vocabulary, the single terms (isolates) of which are then examined and assigned according to five fundamental categories (facets):

Personality	Things or kinds of thing (primary facet)
Matter	Materials
Energy	Operations, activities, actions
Space	Place
Time	Time Period

Ranganathan asserted that these five Fundamental Categories could be used to determine and analyse the components of any subject but they would not necessarily all apply to every subject. Like his enumerative predecessors, Ranganathan still maintained a set of Main Classes but, unlike enumerative schemes, each main class contained only single terms; complex subjects were not represented. Thus, the onus is on the classifier to determine the components embodied by the bibliographic item and then 'build', or synthesise, a representative construction using these separate elements, famously likened by Ranganthan to the building blocks of a Mecanno set. Faceted classifications such as Colon, therefore, are not predicated upon ready-made constructions and are thus much shorter in length than enumerative schemes, taking full advantage of the economies of term reuse. Moreover, as Svenious (1992, p.180) observes, they provide a flexibility, specificity and hospitality to accommodate advances in knowledge, superior to traditional 19th century enumerations:

> ...Ranganathan worried over the inability of enumerative classifications to keep pace with knowledge; in particular, he deplored their lack of expressive power in providing coextensive numbers for subjects. Enumeration is uneconomical as is clear if one imagines trying to enumerate all the sentences in a language like English. It is much more economical to postulate categories of terms, such as nouns, verbs etc., along with rules to combine terms belonging to these categories into sentences.

Ranganathan was a catalyst for change. He pioneered an alterative dynamic and multidimensional view of the universe of knowledge and how this should be analysed and represented within a classification scheme. So influential were Ranganathan's ideas that, in 1955, a group of British classification experts, the Classification Research Group, pronounced that his faceted classification should form 'the basis of all methods of information retrieval'.

The Classification Research Group (CRG) was formed in 1952, emanating from discussions following a Royal Society conference on Scientific Information in 1948 (Classification Research Group, 1955). At this time, the Second World War had generated increased scientific research activity, which was now freed from wartime security restrictions. The resultant influx of research reports needed organisation yet there were concerns amongst scientists as to whether current methods would suffice (Cleverdon, 1997). As a result, the CRG set out to "review the basic principles of bibliographic classification, unhampered by allegiance to any particular published

scheme" (Classification Research Group, 1955, p.1). Despite this initial broadminded aim, the CRG subsequently arrived at the conclusion that every system of information retrieval should be based upon classification and, more specifically, that in order for all semantic relations to be correctly and explicitly represented, this should take a faceted approach (Classification Research Group, 1955). As previously stated, the CRG was strongly influenced by Ranganathan's ideas, brought to Great Britain in 1951 by Palmer and Wells (Beghtol, 1995, p.201). However, they believed that Ranganathan had over-generalised with just five fundamental categories and that, although these could form a basis for analysis, each subject should be examined in its own right, allowing for further categories. Unlike Ranganathan, the CRG initially tested their ideas by concentrating upon discrete subject areas, developing several 'special' as opposed to one 'universal' classification[2], leading them to extend Ranganathan's categories from five to thirteen. However, they maintained that these thirteen fundamental or 'standard' categories may not be applicable nor exhaustive in every subject area (Vickery, 1960, p.24), yet, as the renowned thesauri creator Alan Gillchrist notes, it was a list that was to prove "useful to compilers of schemes ever since" (Gillchrist, 1994, p.104). Moreover, to aid the classifier, the CRG allowed some common compounds to be pre-coordinated within their faceted classifications whereas Ranganathan permitted only simple terms to be listed, creating an 'adultered' as opposed to a 'pure' faceted classification (Broughton, 2001b). Thus, through practical application, the CRG established a procedure of facet analysis for special classifications, which is briefly outlined below (Vickery, 1960; Broughton, 2001a):

---

[2] Citing the Classification Research Group, Foskett (2000) makes a distinction between a 'general' classification, being a *series* of separate classifications aimed at covering a large proportion of the universe of knowledge (the initial 'special' classification approach of the CRG), in contrast to a 'universal' classification, which covers the entire universe of knowledge within the *one* scheme (e.g. the CRG's later focus on the Bliss Bibliographic Classification, to be discussed). It is noted, however, that other writers have used the term 'general' to describe Foskett's 'universal' classification; hence the two terms are often used synonymously.

1. ***Select a homogenous subject area, e.g. office management, diamond technology, food technology etc.***

2. ***Detailed analysis of the literature in the subject area to obtain terms, ensuring that the classification is based upon literary warrant.***

3. ***Group this vocabulary into mutually exclusive categories, according to function and a single characteristic, using the Standard Categories as a basis.***

4. ***Deconstruct any compounds into simple terms, where appropriate, and identify synonyms***

5. ***Within each basic category (facet), group its content terms into subclasses by a single principle of division (order in array) followed by further subclasses (a chain), as appropriate. The terms within arrays can be displayed hierarchically.***

6. ***Assign an order within each facet (e.g. alphabetical, abstract to concrete, simple to complex etc.) appropriate to literary warrant and/or user needs, and list the facets (Schedule Order/Citation Order) from, for example, abstract to concrete or general to special. This creates a reverse filing order of increasing specificity on the library shelves (Principle of Inversion).***

7. ***Assign notation***

Figure 2.3: Procedure of Facet Analysis

As Figure 2.3 illustrates, unlike enumerative schemes such as Dewey, the notation is the final consideration not the driving force. However, despite being a faceted scheme, note that the hierarchical structure characteristic of enumerative schemes is still considered a useful structuring device within facets. The CRG documented this theory of facet analysis, and that of classification in general, in several seminal publications during the 1950s to 1970s, many of which are classic textbooks still today (Mills, 1960; Foskett, 1996[3]; Vickery, 1960; Maltby[4], 1975). During this time, there was an increased attention to theoretical development yet, as previously, always grounded in practice. The impetus was perhaps the focal point of the first International Study Conference on Classification Research held at Dorking in 1957, when several now eminent classificationists, including Ranganathan and members of the CRG, convened to inaugurate classification as worthy of research. It was here that the CRG presented, and gained agreement, that 'faceted classification was the basis

---

[3] Now in its fifth edition, A.C Foskett's 'The subject approach to information' was first published in 1969.

[4] In 1975, the fifth edition of W. C. Berwick Sayers' original 1926 work, *A Manual of Classification.*

for all methods of retrieval' (Classification Research Group, 1955) and, according to Foskett (1997, p.ix), "...the renaissance of H.E. Bliss's great Bibliographic Classification could be said to have begun...".

Henry Evelyn Bliss first published an outline to his 'Bibliographic Classification' (BC) in 1935, with a completed full edition not then appearing until 1953 (Sayers, 1975). The BC began life as an enumerative scheme but was notable particularly for the scholarly basis and logic of its Main Class order (Bliss Classification Association, 2004; Broughton, 2001b). As Sayers (Sayers, 1954 cited in Campbell, 1977, p.4) notes of Bliss's extensive research in establishing an order for his classification:

> ...an unusually exhaustive examination of the ways in which men in their social, philosophical, scientific, professional and industrial and educational relations have organized things. Several orders of knowledge were reached both by induction and deduction. On these, but particularly on the educational and scientific consensus revealed, Bliss rested his classification order.

This Main Class 'consensus' order was maintained in 1967 when members of the Classification Research Group formed the Bliss Classification Association, under the direction of Jack Mills, and sought to revise BC in accordance with their newly emerging principles of facet analysis. Under the guise of the Bliss Classification Association, members of the Classification Research Group were therefore given a universal, as opposed to the previous special, classificatory platform to develop their ideas. The Bliss Classification Association (2004) claim that their subsequent "revision has been so radical that it is more accurately described as a completely new system." The new version of the Bibliographic Classification, known as BC2, was first published in 1977 (Bliss Classification Association, 2004) and employed the Classification Research Group's thirteen standard facet analysis categories and established a standard citation order for these facets, whilst still maintaining the flexibility of Bliss's pioneering idea of alternative location[5] by providing orders alternative to the 'standard', depending on a library's chosen primary facet. The role and compromises of a citation order have been alluded to briefly in this chapter, in so much as the primary (initial) facet within a facet citation order will be collocated at the expense of distributing the secondary facet(s) across the classification schedule,

---

[5] Bliss recognised and provided for topics that might not fit neatly into the 'consensus order', by allowing some concepts alternative main class locations. For example, a library could choose to locate photography under either 'Technology' or 'Arts', but could not choose both as then scatter, as opposed to collocation of like with like, would occur (Maltby, 1975).

but will be explored in depth in Chapter 3. Gillchrist (1994, p.90) argues that the work published during the revision of the Bliss Bibliographic Classification was, in the 20th century, "the most successful attempt to challenge the hegemony of the DDC and the UDC." Moreover, as Broughton (2001b) notes, it was also hoped to be a culmination of the ideas and theory of the CRG, embodied within one universal classification scheme. However, despite being designed as a universal faceted classification, it is perhaps ironic that, today, BC2 is used predominately by small university college libraries (for example, Sydney Sussex College, Cambridge University) and special libraries. This disappointingly limited adoption of the faceted BC2 was not new; commentators observe that H.E Bliss himself suffered a similar apathetic reception:

> Mr. Bliss was a true scholar. His goals and aspirations were different from those of Melvil Dewey, whom he certainly surpassed in intellectual ability and drive. Dewey was a businessman, but he was in no sense as profound in his accomplishments. It is true that Dewey's classification filled a terrible void. As Mr Bliss said, if it had come a few years later the entire course of classification history might have been changed. Many years later, when the Library of Congress was acutely aware of the shortcomings of Dewey for the research scholar, Mr. Bliss was on the scene: but here again he lacked what Dewey possessed, and failed to press the issue of his own system.
>
> (Eugene Garfield 1974, cited in Campbell, 1977, p.6).

Similarly, Maltby (1975, p.218) praises the scheme as having "many merits and few faults" but laments its lack of support, even so shortly after its inception:

> Mills once wrote of a scheme being partly, at least, a test-bed on which ideas and theories could be hammered out rather than a widely used system; he was speaking of CC, but - whatever its reception - these words would seem to fit the revised Bibliographic Classification admirably...It deserves both careful study and greatly increased application; it will certainly have one, but receipt of the other is more dubious and problematical and will not come about without appropriate 'advertising' and marketing campaigns.
>
> Maltby (1975, p.218-219)

Due to the domination of established schemes such as Dewey, LCSH and even UDC, this 'marketing' drive still remains a topical issue, as evidenced in an Annual Report (2001) on the Bliss Classification Group homepage, which contains discussion relating to 'Publicity and Increasing Membership' (Bliss Classification Association, 2002). The latest Annual Report (Bliss Classification Association, 2003) details efforts to promote the scheme, including the production of leaflets, management of the website and organisation of short courses. Despite such initiatives and regular

updates to the BC2 Schedules via the *Bliss Classification Bulletin*, a classification scheme can only survive with the size and support of its user base.

Thus, the historical entrenchment of 19th enumerative/hierarchical schemes still prevailed, despite the ingenuity of Ranaganathan in instigating a new treatment of complex subjects via faceted classification, coupled with related initiatives from the Classification Research Group to devise a new universal (i.e. Bliss) and several special faceted classifications. Although such faceted schemes offered enormous potential and flexibility to meet the needs of an increasingly complex information landscape and its users, the historical legacy of established schemes precluded their adoption both on their inception and still, more pertinently, today. This may be due to a number of factors. Firstly, depending on the desired purpose of the classification scheme within a particular library, the financial implications of converting a library collection might outweigh the perceived benefits. Secondly, the now common practice of sharing bibliographic MARC[6] records amongst libraries, containing recommended class numbers, favours the adoption of a 'standard' classification. Thirdly, Cataloguing-in-Publication records (CIP) supplied by Bibliographic Data Services Limited (no date) contain Dewey and, for academic titles only, Library of Congress Classification codes. Fourthly, many libraries now purchase 'shelf-ready' books with pre-assigned classification codes, supplied from an outside agency, who are perhaps reluctant to support the less widely used schemes. Moreover, as mentioned previously, the perceived 'unnecessary' complexity of synthesising classificatory components and the resultant notation, particularly in the case of the Colon Classification and even the Universal Decimal Classification Scheme (UDC), may also have overwhelmed or seemed superfluous to librarians and end-users. Thus, the heavily disproportionate usage in favour of the older enumerative/hierarchical schemes over the newer faceted schemes still remains to this day. As stated previously, though, it could perhaps be seen as recompense that despite not being the panacea envisaged by the Classification Research Group, faceted schemes have had a lasting effect upon classificatory theory, which has infiltrated into the practice of popular, traditionally enumerative/hierarchical schemes. Thus, libraries may not have embraced faceted classification directly and whole-heartedly but its influence is still

there indirectly, in the form of an extension to the more established, chiefly enumerative predecessors, notably the Dewey Decimal Classification Scheme.

The influence of faceted classification upon the DDC is today clearly apparent; however, as Miksa (1998) and Sweeney (1983, 1990) note, such influence was relatively late in the making. Although, as stated previously, Melvil Dewey included recognised recurrent categories such as form, geographic area, genre and language as early as his first edition in 1876, Miksa (1998) notes that these were not listed as separate auxiliary tables with instructions to synthesise with topical divisions (as they are today), but were enumerated, and hence repeated, across relevant classes. Therefore, although recognition of recurrent concepts was apparent, the principle of concept re-use was not employed. Despite this early awareness of recurrent categories, the enumerative approach then prevailed for over fifty years until 1932, when, influenced by the UDC, the 13th edition saw recurrent categories ('Common Subdivisions') listed only once in a separate auxiliary table, and then, more explicitly in 1965 (17th edition), when the term 'facet' was formally recognised for the first time (Miksa, 1998). The subsequent 18th edition in 1972 then saw the creation of all seven auxiliary tables (now in the 22nd edition reduced to six, due to the removal of Table 7, replaced by re-use of Schedule notations and notations in Table 1). These auxiliary tables can be used to classify materials with complex subject matter by synthesising classificatory components to create a notation. Table 1 (Standard Subdivisions of recurrent non-subject related forms) can be appended to any Dewey Schedule component both with and without instruction. The remaining six auxiliary tables are as follows and are only added under instruction within the Schedules or other Tables (Chan et al, 1996, p13):

---

[6] **MA**chine **R**eadable **C**atalogue records are in a structured and standardised format to enable manipulation and exchange by computer.

Table 2 Geographic Areas, Historical Periods, Persons
Table 3 Subdivisions for the Arts, for Individual Literatures, for Specific Literary Forms
Table 3A Subdivisions for Works by or about Individual Authors
Table 3B Subdivisions for Works by or about More than One Author
Table 3C Notation to be added where instructed in Table 3B, 700.4, 791.4, 808-809
Table 4 Subdivisions of Individual Languages and Language Families
Table 5 Racial, Ethnic, National Groups
Table 6 Languages
Table 7 Groups of Persons

These Table components and notations, therefore, combine with the Schedules (Main Classes and their subdivisions), previously detailed in Section 2.4. In addition to this method, Dewey notations can also be synthesised ('built') under instruction by re-using Schedule components or in-built Schedule tables, which apply to a specified range of Dewey codes. Thus, although the introduction of these synthetic devices has increased the level of complexity when assigning classificatory notations, it has succeeded in creating a new depth of analysis and representation, albeit within a scheme with a still predominantly enumerative and hierarchical basis. However, as will be discussed in successive chapters, challenges created by this historical legacy still remain. As stated at the beginning of this section, the recognition and incorporation within Dewey of the specificity and flexibility provided for by the synthesis and faceted principles of 20th century theorists was relatively late, even reluctant. Prior to the introduction of the seven auxiliary tables, the DDC had been criticised for its lack of detail, exacerbated by the controversial Standard Edition (15th edition) in 1951.

The idea of a Standard Edition arose from the uneven levels of specification, exposed when classes were compared; some classes were extremely deep with many hierarchical levels whilst others had relatively few (Miksa, 1998). A Standard edition aimed at rectifying this, either by extending those lacking depth or by reducing those that were perceived as too deep. Unfortunately, the latter was preferred:

> The 15th edition virtually shocked the library community because if the drastic nature of the changes it represented. The first major change was to limit subject specification in the system to between four and six levels of hierarchy. This meant limiting class numbers to six digits (with very few slipping through at seven), although many were only three to five digits in length. The effect of this change alone was to radically reduce the system in size. One way to measure the change is in page counts. The schedules were reduced from approximately 1,050 pages in the 14th edition to only 467 pages in the 15th.
>
> Miksa (1998, p.21)

Such was the adverse response that the edition was revised and reissued with uncharacteristic haste in 1953 and many academic libraries even began to convert to the Library of Congress Classification Scheme (Miksa, 1998). It has been argued that this lack of appreciation for specificity, and the subsequent outcry, arose from the differing views between the United States and Great Britain regarding the purpose of classification (Miksa, 1998; Sweeney, 1983, 1990). In the United States, a 'dictionary' author A-Z library catalogue was preferred and classification was solely a 'mark and park' shelving tool; whereas Great Britain had preference for a classified catalogue and was also therefore utilising classification for information retrieval, where specificity is paramount. These different approaches will be explored further in the following Chapter 3. In our discussion of classification here, though, it is worthy of note that this perhaps delayed the acceptance and incorporation within Dewey of the primarily European developed principles of faceted classification. Such was the disenchantment with Dewey's lack of detail and provision for complex subjects that librarians producing the British National Bibliography (BNB), the largest Dewey classified catalogue of its time, began to create their own "...faceted schedules completely alien in both structure and in notation to the Decimal Classification" (Sweeney, 1983, p.195). Despite an increased capacity for building compounds and the introduction of a separate area table, as stated above, the BNB editors rejected both the 16th (1958) and 17th (1965) editions. It was only in 1970 that the BNB relinquished its own faceted schedules, adopting the DDC wholeheartedly. However, interestingly, Sweeney (1983, p.197) believes that " the decision was more to do with the introduction of machine-readable bibliographic records than with the desire for a coherent classification scheme." Thus, as stated previously, the potential (or even pressure) to standardise prevails over the 'tailor-made'. Possibly as recognition of this long-awaited BNB endorsement, also in 1970, the Dewey Editorial Policy Committee (EPC) invited the Chairman of the reconvened Library Association Dewey Decimal Classification Committee (LADDC), Joel Downing, to attend the EPC provisionally for one year, with subsequent renewals in later years (Trotter, 1993). This representation provided a direct British influence on the development of the DDC, which had previously been restricted to North American librarians, resulting in a greater influence of faceted classification; for example, the creation of a considerably revised Music Schedule in 1973, based on the faceted principles of the Classification

Research Group's, principally Eric J, Coates', British Catalogue of Music Classification (Trotter, 1993). Today, as exemplified in section 2.4, although the tendency for an American bias is still apparent in Dewey's structure and captions, British (and now international) representation and consultation is more established and assured via liaison between the EPC and, in Britain, the converged and renamed British professional body, which succeeded the Library Association: the Chartered Institute of Library and Information Professionals (CILIP) and its Dewey Decimal Committee.

The above discussion of twentieth century classification has illustrated how the burgeoning information context and the growing demands of scholars and scientists led to a more specialised and flexible approach towards classification, providing alternatives to nineteenth century mass enumerations and promoting theoretical development. Influential in theory rather than revolutionary in practice, such twentieth century synthetic and faceted classifications heralded recognition and accommodation of complex subjects. The twenty-first century, however, is posing further challenges for knowledge organisation and representation.

2.7 Twenty-first Century Classification: Interdisciplinary Challenges

The late twentieth and early twenty-first centuries have seen a fragmentation of traditional subject discipline boundaries (Williamson, 1998); subjects are studied topically, across disciplines, yet the most widely used hierarchical/enumerative classification schemes, such as the DDC, still organise on a disciplinary basis:

> The major systems that have predominated in the twentieth century were originally predicated on the academic disciplines. This structural principle is no longer adequate because multidisciplinary knowledge production has overtaken more traditional disciplinary perspectives and produced communities of cooperation whose documents cannot be accommodated in a disciplinary structure.
>
> Beghtol (1998b, p.1)

Thus, the challenge of the preceding twentieth century environment, to represent items having complex subjects where the various components fall within the *one* subject discipline, has now perhaps been superseded by the task of accommodating

those where the components span two or more subject disciplines: the phenomenon of interdisciplinarity.

Since disciplinary classification is incongruently organising an interdisciplinary universe of knowledge, those scholars who study across traditional disciplinary boundaries may find their subject scattered throughout the classification and, in turn, the library building itself. These changes in knowledge production have prompted researchers (Beghtol, 1998a and 1998b; Klein, 1996; Williamson, 1998; Bates, 1996; Searing, 1996) to examine the information seeking of such interdisciplinary scholars and how resources can be best classified and organised to meet their needs. Various possible solutions are being proposed from the unlikely creation of entirely new universal schemes with a non-disciplinary basis, through to incorporating faceted principles as extensions to established universal classification schemes to examine topics by phenomena rather than purely discipline, and, more radically, the provision of multiple access points, not physically on the library shelves but electronically within a library catalogue or other information retrieval tool:

> In the electronic environment, texts are freed from the limitations of physicality, so in theory, any number of classes can be assigned.
>
> Searing (1996, p.320)

As such, the twenty-first century has provided the potential for classification to move beyond its traditional remit of physical access, grouping related items on shelves according to one physical location, to a context in which such restrictions do not necessarily apply and a number of access points can be provided for. Realising the perceived potential and exploiting the immense intellectual effort of traditional classification schemes within an electronic, not physical, environment is central to this current PhD research. However, the following chapter suggests that the degree to which classification is currently being recognised and utilised in this additional role as an electronic information retrieval tool is relatively low in comparison to other retrieval methods. Indeed, where classificatory structures are employed as a means of browsing a physical or even virtual collection, these applications adhere to a physical access 'model', unnecessarily replicating the confines of the physical world.

2.8 Conclusion

This chapter has charted the development of classification, as a tool for the organisation and retrieval of information, from its ancient beginnings to the present day. Societal influences have always exerted an impact upon the need for and subsequent development of schemes: from simple physical 'separations' to imbue some sense of order to early records of knowledge; followed by the desire to map the universe of knowledge in more elaborate philosophical and scientific classifications, together with the parallel pragmatic efforts of booksellers from which modern day library classification began to emerge; through to the development of several universal classification schemes, which are still used today.

It has been demonstrated that the development of these universal schemes began with the Dewey Decimal Classification in the late nineteenth century, coinciding with the emergence of open access libraries. With a classificatory structure arguably influenced by the preceding philosophical/scientific classifications, these nineteenth century enumerative and hierarchical schemes still predominate today, despite efforts (motivated by a specialisation in knowledge during the early twentieth century) to provide for more specificity, flexibility and hospitality for complex subjects. Although not the panacea that was once imagined, the advantages of facet analysis, coupled with its new conception of the universe of knowledge, was not revolutionary in practice due to the continuous entrenchment of traditional nineteenth century schemes.

Although widely used nineteenth century schemes have adopted certain principles of faceted classification, the challenges of the twenty-first century call for yet more flexibility to accommodate the interdisciplinary fragmentation of knowledge and the increasingly diverse needs of users. This may not be achievable within the physical domain of shelf classification but the electronic environment provides the potential for additional organisational and retrieval tools, allowing classification to be manipulated to provide alternative arrangements of a physical collection. This tool is the Online Public Access Catalogue, which is the focus of the following chapter.

CHAPTER 3

CLASSIFICATION, LIBRARY CATALOGUES AND THE PROBLEM OF SUBJECT ACCESS

3.1 Introduction

In parallel to the previously discussed development of classification systems, the growth and diversification of knowledge and, in turn, its embodiment in library collections led to the need for an additional retrieval tool: the library catalogue. Certain types of early catalogues were founded upon foregoing classificatory structures of physical organisation, indicating hierarchy and conceptual relationships; however, the emergence of computerised systems and the Online Public Access Catalogue (OPAC) heralded a different retrieval paradigm, one of 'exact-match', in which classification was to play a lesser role.

In contrast to other information retrieval systems, the OPAC interface and search interaction has changed little since its inception. Although suitable for searches when the author and title are known, the 'exact-match' query paradigm causes problems for more exploratory subject searches. The re-discovery and re-orientation of classification within catalogues, bringing classificatory structures to the interface, could move current OPACs beyond an 'exact-match' finding tool to one of exploration and discovery, providing this application is freed from the linear conventions of physical access.

This chapter argues that, from printed to online access, the diminishing role of classification within library catalogues has effectively relegated classification from information retrieval tool to shelf-location device, helping create an enduring OPAC model based on finding not discovery and causing problems for certain users when searching by subject. Firstly, these problems are traced back to their origins by defining types of early printed card catalogues and the role, and subsequent demise, of classification within such catalogues, a demise which was to have a future impact upon online catalogue design. Secondly, it is described how a 'standard' specification-based

OPAC interface and interaction arose from the overriding printed dictionary catalogue model, coupled with the influential 'exact-match' information retrieval paradigm, despite challenge from the more recent user-oriented paradigm. Thirdly, the related and consequential problems of subject searching in OPACs are considered and the contributory role of the OPAC interface explored. Finally, it is asserted that whilst the potential value of classification at the system interface for encouraging exploratory subject searching unfortunately remains untapped within commercial library OPACs, it is being discovered and re-cast by the Internet and Semantic Web communities.

3.2 Early Catalogues

A classified physical arrangement, although aiding information retrieval and collocating related items, offers just one access point per library item. As collections grew, libraries recognised a need to document and describe the contents of their collections and to provide additional access points and arrangements (such as by author, title and subject). Using such descriptions and access points, a library catalogue is a searchable repository, holding surrogates of library items in the form of bibliographic records, organised according to a set of standards.

It is thought that the earliest catalogues emerged in the Near East, circa 2000 B.C., in the form of clay tablets to document religious works and, later, in the Alexandrian Library, in the form of 'pinakes' (wooden boards) (Carpenter, 1994). In contrast to the brief listings of their Western European counterparts, such catalogues showed an early awareness of a need to formalise collection descriptions, identifying certain practices for the form of author names and the recording of alternative versions of a text (ibid.). The first printed catalogue appeared in 1595, documenting the library at the University of Leiden. Subsequently, across the globe, libraries began to document their own collections in their own way and, as a result, varying cataloguing 'rules' emerged. In 1876, Charles A Cutter (Cutter, 1904, p.12) attempted to rationalise this diversity by identifying the types of catalogue produced to date and setting out the objectives of the catalogue, which still remain the essence of catalogues today (Smith, 1997):

OBJECTS
To enable a person to find a book of which either
(A) The author
(B) The title is known
(C) The subject
To show what the library has
(D) by a given author
(E) on a given subject
(F) in a given literature
To assist in the choice of a book
(G) as to its edition (bibliographically)
(H) as to its character (literary or topical)

Although both attempting to fulfil Cutter's 'objects', yet in different forms and with different ramifications for the role of classification, two main types of catalogues had been established: the dictionary catalogue and the classified catalogue.

Before the advent of the computer, classification was utilised not just for shelf arrangement of a library collection but also for information retrieval within a classified catalogue. Such catalogues provided a systematic arrangement of library item surrogates (bibliographic records) by the classification number, hence creating a browsable file of related item records, as they would appear on the library shelves. This classified file was then supplemented with a separate alphabetical subject index file and an author surname file, to provide additional access points. In contrast, the dictionary catalogue, or alphabetical catalogue, comprised just one single A-Z file of author names, titles and subjects, although Smith (1997) notes that larger libraries often preferred to split this file into two sequences, by author/title and subject. The arrangement of the dictionary catalogue tends to favour the library user who knows exactly what they are looking for, as reflected in early observations by Sharp (1948, p.20):

> It appeals particularly to people who want very specific information, or who are not pursuing a detailed investigation of a wide branch of knowledge. The gardener who wants books about roses only, or the woman who is interested in leatherwork and in no other handicraft, can be served simply, yet partially, by this catalogue...

Thus, in a dictionary catalogue such users would need to search under 'R' for roses and 'L' for leatherwork, an alphabetical organisation, which will inevitably scatter items on related subjects, as Sharp (1948, pp.20-21), again, clearly illustrates:

> A reader going to Italy will certainly want to see what books the library has on Rome, but he may be going to Florence, Venice and Naples as well; he will need to see some general guides on Italy too. If he is using a dictionary catalogue, it will be necessary to consult all these headings - and possibly others - in widely separated parts of the catalogue before he can ascertain the library's complete resources on all these places.

Conversely, the classified catalogue collocates items on related subjects, although as will be demonstrated in Chapter 4, if considered according to the differing needs and perspectives of users, in some cases the classification scheme *per se* can in fact be seen as *creating* scatter. It would be naive to assume that a single classified arrangement can meet the needs of all users, but the classified catalogue built upon the intellectual effort applied at the library shelves can provide more enhanced subject retrieval than its dictionary catalogue counterpart. However, although favoured in Europe, the classified catalogue was "virtually unknown" in the United States (Sweeney, 1990, p.42).

In the card catalogue era, there was a preference for the classified catalogue in European libraries yet the dictionary catalogue predominated in the United States (Sweeney, 1990; Smith, 1997). As discussed in the previous Chapter 2, this led to divergent views between these two continents as to the perceived role and development of classification: whilst the United States viewed classification solely as a 'mark and park' shelf organisation tool, Europe also utilised classification for information retrieval in their library catalogues. As discussed, just as historical entrenchment, standardisation and bibliographic record sharing seemed to discourage the wholehearted adoption of newer faceted classification schemes by libraries, similar factors seemed to play a role in the corresponding demise of the classified catalogue.

The year 1901 saw the beginnings of cooperative cataloguing and a union catalogue, when the American Library of Congress began to sell its catalogue cards to other libraries and accept card exchanges (Carpenter, 1994). As Lerner (2000, p.80) states:

> Librarians knew there was no need for the same book to be catalogued in detail by every library that owned it. After all, the basic bibliographical details of the ordinary printed book – author, title, imprint, physical aspects – did not vary from one copy to another. Nor did its content.

The efficiency, consistency and quality control of such shared cataloguing was extremely appealing to libraries and forms the cornerstone of cataloguing practice today. The

Library of Congress Catalogue, however, was a dictionary catalogue, based on the aforementioned Cutter's *Rules for a Printed Dictionary Catalogue* (Cutter, 1904; Carpenter, 1994); therefore, the standards deriving from catalogue cards intended for a dictionary filling order, by default, were receiving a larger audience than those of the classified catalogue. This catalogue card exchange service soon led to the establishment of the 'National Union Catalogue' and a 'national bibliographic service'. Meanwhile, in addition to the establishment of both the dictionary and classified catalogue filing orders, international standards for the actual format of the bibliographic record descriptions and access points were beginning to be explored and agreed.

In the nineteenth century, as previously mentioned, different countries had already begun to create their own cataloguing principles, most famously Panizzi's pioneering '91 rules' for the British Museum catalogue (1841) (Joint Steering Committee, 2004). Prompted by Melvil Dewey, cooperation between Britain and the United States began in 1904, leading to the first 'international' cataloguing code, published in both American and British editions, in 1908 (Joint Steering Committee, 2004). However, the advent of the Second World War saw an end to this initial cooperation and the American Library Association and Library of Congress were left to continue the development alone, culminating in the 1949 *ALA Cataloguing Rules for Author and Title Entries* and the *Rules for Descriptive Cataloguing in the Library of Congress (*Joint Steering Committee, 2004). More widespread, truly international, agreement amongst increasing diversity was sought, leading to the 1961 International Conference on Cataloguing Principles and its resultant 'Paris Principles'. Cooperation between Britain and North America (Canada and United States) began again and, under the direction of the Library of Congress, the above standards all fed into the development of British and North American versions of the *Anglo-American Cataloguing Rules (AACR)* in 1967, a single version of its second edition in 1978 (AACR2) and, today, its latest 2002 revision. Due to the North American impetus and the stronghold of the Library of Congress, such rules and standardisation emanated from the dictionary, not classified, catalogue. Moreover, with the development of the Machine Readable Cataloguing in the early 1960s being based on Library of Congress catalogue records (with subject access provided by Library of Congress Subject

Headings not classification) and then the first online catalogues being developed in the United States, the dictionary catalogue successively became the model for the Online Public Access Catalogue (Smith, 1997; Carpenter, 1994). With the demise of the classified catalogued and its principles, the role of classification as an information retrieval tool diminished, confined to a 'mark and park' shelf location device. Langridge (1997, p.x) laments classification's, namely faceted classification's, relegation from being hailed at Dorking in 1957 as "the basis for all methods of information retrieval":

> By the end of the sixties this latter idea had been swept away by the misguided preoccupation with computers. The steady flow of books on classification theory dried up. Organisations that badly needed more effective subject retrieval preferred spending money on computerising their catalogues to improving their classification.

And, so, the information retrieval paradigm shifted.

3.3 Early Online Public Access Catalogues and the Information Retrieval System Legacy

Library Online Public Access Catalogues (OPACs) are just one subset of information retrieval systems that were undergoing parallel development. Information retrieval arrived with the advent of computer technology in the 1950s. Early systems of the 1950s and 1960s were batch-mode systems, which entailed a single 'one shot' result set and, apart from the initial query, no other user-system interaction (Harter and Hert, 1997). System interfaces were not intended for end-users but highly trained intermediaries who were required to master an often complex command set. Similarly, computerised library systems did not begin life as tools for the end-user. Yee and Layne (1998) note that the first (usually experimental) library systems of the 1960s were used to automate item circulation and other library processes. The 1970s saw the emergence of commercial systems, but again not for public access. It was not until the early 1980s that systems designed for the end-user (Online Public Access Catalogues) were produced. Hildreth (1995) argues that early OPACs were designed as either an attempt to replicate the functionality of the card catalogue, or following the model established by commercial online information retrieval systems. However, in either case, as such early systems were generally not made available to untrained end-users, interface development was not a concern.

Early OPAC interfaces allowed searches only by author or title, influenced by the exact match query-oriented paradigm of commercial information retrieval systems. Designed following an expert user intermediary model, such systems depended on the user knowing the initial part of a library item title or author name and exactly matching this to the unseen vocabulary of the system. Online Public Access Catalogues, therefore, were borne from a 'known-item' dictionary catalogue search model with, initially, no subject search provision. Moreover, information retrieval, of which OPACs are a part, was emerging as a scientific discipline with focus on providing technical solutions, yet evaluated using performance metrics that excluded users. A comprehensive review of information retrieval research is beyond the scope of this thesis, but since it provides a context to past and present OPAC interfaces and evidence as to the limitations of their design (discussed in the following section 3.5), a brief outline of themes is given below.

3.3.1 Computer-Orientated Information Retrieval

Early information retrieval system evaluations saw the development of metrics to ascertain the quality of the result set, not the actual information search process. Originating from tests on the Uniterm system at the Cranfield College of Aeronautics in 1953 (Rowley and Slack, 1998), the concept of 'relevance' and the related measures of 'recall' and 'precision' are well documented in the information science literature. Such metrics are based on establishing a query and accompanying relevant document set in advance of the actual search interaction, and then evaluating the success of this interaction in terms of the extent to which the system retrieved relevant documents from the pre-determined relevant document set (recall) and the extent to which the pre-determined relevant documents were retrieved from the total document set (precision). Such metrics, therefore, do not relate to the user but, instead, are based on search output, the product not the process. In their 'Modern Information Retrieval' textbook, Baeza-Yates and Ribeiro-Neto (1999, p.7) distinguish two approaches towards information retrieval:

> "...a computer-centered one and a human-centered one. In the computer-centered view, the IR problem consists mainly of building up efficient indexes, processing user queries with high performance, and developing ranking algorithms which improve the 'quality' of the answer set. In

> the human-centered view, the IR problem consists mainly of studying the behavior of the user, of understanding his main needs, and of determining how such understanding affects the organization and operation of the retrieval system...we focus mainly on the computer-centered view of the IR problem because it continues to be dominant in the market place."

Although more recent computational information retrieval techniques and evaluation has incorporated the user by way of relevance judgements, interactive query expansion and relevance feedback (see Efthimiadis, 1994), the focus of this approach is still very much upon studying and improving information retrieval techniques and not the users' needs, subjective reaction towards an information retrieval system and role within the search process. Despite Baeza-Yates and Ribeiro-Neto's comment that the computational approach is "dominant in the market place", since the late 1970s there has been a growing and now very strong interest in studying user behaviour, information seeking and the information search process.

3.3.2 User-Orientated Information Retrieval

In contrast to the computer-oriented information retrieval paradigm, user-oriented research has considered the user's cognitive state and, moreover, the search process rather than product, presenting taxonomies that emphasise the dimensions and continuum of information seeking. Focus upon the user and the information seeking process challenged early assumptions behind information retrieval system and OPAC design: that users have a clear and unchanging information need, which they are able to formulate into a precise query to match and retrieve a single ideal set of documents represented by the search terms. However, research has shown that users initially have vague information needs and difficulties in formulating such queries. The beginnings of research to overturn this traditional assumption were seen in a seminal paper by Taylor (1968), which studied user-librarian (not user-computer) interactions at the reference desk. Despite only interviewing librarians, not also users, about the reference interaction, Taylor (1968, p.179) promoted the view that such interactions were a changeable process:

> ...an inquiry is looked upon not as a command, as in conventional search strategy, but rather as a description of an area of doubt in which the question is open-ended, negotiable and dynamic.

Users were not viewed as having a single, clearly defined information need but a series of needs, affected by the "constraints of the system and its files" (Taylor, 1968, p.183):

> Q1 – the actual, but unexpressed need for information (the visceral need);
> Q2 – the conscious, within-brain description of the need (the conscious need);
> Q3 – the formal statement of the need (the formalized need);
> Q4 – the question as presented to the information system (the compromised need).

According the Taylor, users therefore begin with a "vague sort of dissatisfaction", which gradually progresses into a clearer conception and expression of the information need, which then, however, is required to be "compromised" according to the vocabulary and capabilities of the information retrieval system. As will be discussed in the following section, this translation from the 'formalised' to 'compromised' need is when problems are often encountered. Although, today, rather a simplistic model of the information seeking process, Taylor's classic paper was extremely influential and, according to Belkin (1990, p.12) and Hewins (1990, p.156) was a "precursor" to the so-called cognitive view of the information seeking process.

Following Taylor's concern with user needs, from the late 1970s, certain researchers (e.g. Belkin et al, 1982; Ingwersen, 1992; Borgman, 1986) began to focus upon modelling the individual's information need and knowledge state in order to inform information retrieval system design (Allen, 1991). The goal was to encode systems with related elements of the user's mental mode in order to increase the compatibility between user and system (Smith et al, 1987). The cognitive viewpoint moved away from a solely 'input-output' traditional view of information retrieval by also considering the information retrieval process. Moreover, unlike the traditional information retrieval view of a clear user need, the cognitive viewpoint considers that the user has an unclear conception of this, the most well-known representation being Belkin, Oddy and Brookes' (1982) idea of an Anomalous State of Knowledge (ASK), motivating the information need and defined by Belkin (1981, p.46) as: "recognition by the user that her/his knowledge of a topic or situation of concern is inadequate." As such, systems were encoded and designed with the requisite knowledge, with a view to reducing such uncertainty. Such designs were

typically based on elicitation and analysis of 'think-aloud' verbal protocols from, or interviews with, expert intermediaries and the individual with the information need. The intermediary's knowledge structures, the 'vocabulary' of the system, then formed the knowledge base of the 'expert' system. However, such knowledge structures were hidden from the user in internal 'frames' or 'scripts', an application aligned with artificial intelligence research, in a question-answer natural language based end-user interface, aiming to re-create and obviate the expert intermediary. Thus, despite recognising that users approached systems with an unclear conception of their information need and assisting the development and focusing of this need via a question-answer interaction, such early cognitively based systems still required the user to interact with a predominantly query-based search interface, motivated by specification, albeit more elaborate, rather than recognition. With the advent of hypertext, direct manipulation, graphical user interfaces and increasing research into the behaviours and process of information-seeking, in his subsequent work, Belkin et al (1993, p.329) recognised that query specification was just one part of a larger, more complex process:

> In this earlier work, we relied on the analysis of the specific situation of user-intermediary interaction in document retrieval situations... This approach was useful in an early development of our CTA [Cognitive Task Analysis] for IR [Information Retrieval], and we were able to focus it specifically on one aspect of interactive IR, that of query formulation and reformulation. However, the context of this situation on which the CTA is based led us to consider behaviours that we now recognise as being in only one general region of our 'ISS' [Information Seeking Strategies] Space - that characterized by *searching, specifying* and *selecting* in *information* resources. Our broader framework of Information Seeking Strategies suggests that we also need to develop a CTA associated with a wider variety of ISSs, such as those associated with scanning, recognizing, and learning in both *information* and *meta-information* resources.

Seminal behavioural research (Ellis, 1989; Bates, 1989) acknowledged that users might engage in several activities during the information seeking process, including the exploratory strategy of browsing.

In an extensive review of the literature, Chang and Rice (1993, p.235) claim that browsing has been subject to various, often seemingly contradictory, definitions, whereby both "planned/unplanned", "goal directed/nongoal directed", "systematic/unsystematic" characterisations, seemingly coexist. Similarly, as Hildreth (1995) notes from his

research review and resultant typology, different types of browsing behaviour can form part of "a family of information seeking activities":

Category 1	Category 2	Category 3
undirected browsing	semi-directed browsing	directed browsing
general browsing	general purposive browsing	specific browsing
serendipity browsing	general purpose browsing	search browsing

Figure 3.1: Browsing Behaviour Categories (Hildreth, 1995)

These typologies comprise three basic levels, distinguished by (a) the extent to which the user has a specific goal in mind and (b) the systematicity of tactics employed (Marchionini, 1995). Category 1 browsing is "largely random, unstructured and undirected"; category 2 constitutes habitual browsing, for example of a particular publication for current awareness; whilst a category 3 browser "has a specific end in mind, but does not approach the catalog with a well-formulated search strategy" (Hildreth, 1995). Thus, browsing can be just as purposeful as the traditional query strategies and should not be disregarded "in contrast to directed searching, to see it as a casual, don't-know-what-I-want behavior that one engages in separately from 'regular' searching" (Bates, 1989, p.414). Despite the increasing study, recognition and complexity of these browsing behaviours, Chang and Rice (1993, p.233) observe that, again, the query-oriented search paradigm tends to dominate:

> Historically, library and information science as well as information systems literatures exhibit a bias toward specific, direct searching as opposed to exploratory, iterative browsing. This bias is partly due to some unrealistic assumptions about users and the nature of information seeking - e.g. that users have unbounded rationality, have static and well-defined information needs, know what they want, and are output oriented.

However, research has shown that the information need and search strategy can evolve as new information sources are encountered (Bates, 1989) and as an individual progresses through a constructive exploration and learning process from feelings of uncertainty to a focus formulation (Kuhlthau, 1993, 1999). Other 'process' researchers (Suchman, 1987; Hert, 1996, 1997; Wilson, 1999; Xie, 2000) have considered information retrieval interactions as similarly varied in strategy whilst situated, not entirely pre-planned but influenced by the search environment, and embedded within a larger information seeking context of different levels of information need, in which the overall task may remain

constant, yet the specific user-system interaction goal may 'shift' dynamically in response to perceived feedback (success/failure) from the current interaction context.[7] Based on inductive, naturalistic and qualitative research, such investigations argue that since the information retrieval interaction is situated, we should move away from providing cognitive models of users or absolute theoretical categorisations of the information search process towards an alternative focus:

> Rather than look for constants, one must instead turn towards developing an understanding of how people use elements from their situation to move and act.
>
> Hert (1997, p.115).

Since the system is an integral part of this 'situation', effective interface design and feedback is thus paramount. As an extension to these situated-action investigations, more recent research is considering a broader view of the social context, beyond the over-arching information seeking task; information is not sought but discovered:

> ...information as something that is constructed by people in their interactions with other people, technology and structures as they move through life and work.
>
> Solomon (2003, p.257).

Influenced by sociological theory and research, this view proposes that individuals 'discover' information as part of their life/work social context and aims to study how this information is constructed to enable the design of effective support systems. Research tends to focus upon studying users' different institutional environments and the information discovery *in situ*.

Therefore, research within the user-oriented paradigm - studying and modelling cognitive processes and appropriate system design; information seeking behaviour and the role of browsing and exploration in clarifying and articulating a vague information need; and the user-system information retrieval interaction and information need as part of a larger information seeking (or discovery) process, affected by time and social context – has all

---

[7] Hert's (1997) findings contradict evidence that the information need changes during the user-OPAC interaction. She discovered that although the search interaction was situated, in contrast to the findings of several researchers whom she cites (viz. Bates, Harter), users' goals remained relatively static. However, she claims that this may be due to a lack of distinction in the literature between the terms 'goal' and 'information need', and suggests that the goal of a specific OPAC search may be part of a larger information seeking process motivated by an information need which may change over time (cf. Kulthau, 1997).

highlighted the challenge faced by information retrieval system designers in supporting such complexity. This research has indicated that users engage in multiple types of searching behaviour and that certain stages of the information seeking process, when users are uncertain and unable to articulate their information need, would benefit from system interfaces that encourage exploration not specification:

> Most systems work fairly well when the problem is well-defined in the collection stage but few are designed for the person who is learning in the exploration stage...can systems be designed that do not close the person down too quickly - that are sufficiently open to accommodate exploring and formulating.
>
> Kulhthau (1999, p.14)

However, as will be discussed in subsequent sections, although experimental information retrieval systems and websites may seek to encourage exploratory information seeking, commercial library OPACs generally still conform to a query-based, usually 'form fill-in', interface, which seems to equate with a more simplistic, or assumed, model of early information retrieval and OPAC research. In contrast to the previously discussed user-oriented 'process' studies, such early OPAC research (see Baker and Lancaster, 1991) has traditionally dichotomised online information seeking into known-item searches and subject searches (often equated with browsing), referring to the specificity of the object sought and the search input, rather than the actual process involved in acquiring that information. However, as research into user-OPAC searching behaviour (Hancock-Beaulieu, 1990, p.335) has suggested, the tool very often "tailors the task" and, similarly, OPAC research itself has followed the model of the tool under investigation, exploring the success of the searches available. Despite this, results reveal the limitations of wholly specification-based OPAC interfaces and relate to the findings of the more process oriented research, namely that users experience problems when they do not have a clear idea of the specific search output. They do not have a library item title or an author (a known item) in mind prior to searching but are seeking information about a particular topic; in other words, they are attempting a subject search.

3.4 The Problem of Subject Access

The problem of subject searching is a finding well documented in the literature (Bates, 1977; Markey, 1989; Larson, 1991a, 1991b; Walker, 1991; Husain and O'Brien, 1992; O'Brien, 1994) and is explored further in preliminary OPAC-user study by the current author (see Appendix 1). Markey (1989) found that "between one-third and one-half of the subject terms that users enter fail to retrieve bibliographic records in the online catalogue" (p.82), with the converse problem in 30-40% of cases of too many hits (Walker, 1991). Common problems cited in the literature include:

- Formulating a query containing the 'right' (matching) subject heading
- Entering search terms which are too specific or too general
- Broadening the search if too little is retrieved
- Typographical errors and misspellings

All of the above are generated by the user experiencing difficulties in expressing and matching his or her information need to the 'unseen' subject vocabulary of the system. This 'unseen' vocabulary and underlying subject access mechanism in the majority of OPACs is currently provided by the controlled vocabulary of the Library of Congress Subject Headings. These headings comprise a string of pre-coordinated index terms. Users can perform a 'subject' search, gaining them entry into the entire alphabetical subject file at the point where their search term appears first in the string. This is often labelled as a 'subject browse'; however with its linear structure, as will be discussed, this is a severe misrepresentation. If users also want to find those items where their search term appears mid-string in the subject index then they must perform a 'subject keyword' search. Whether users understand and appreciate the differences between these two types of subject search is doubtful. Perhaps due to inadequacy of this, studies have shown that users often perform 'Title keyword', 'Title' and even 'author' searches when they are actually seeking subject information (Hancock-Beaulieu, 1990). Lancaster and Baker (1991, p.209) cite several studies that highlight the extent of the problem:

> The initial failure experienced by a patron when trying to match a query term to a catalog subject heading would not be so important if the patron would then try other subject headings. A few highly motivated patrons will actually do so... The typical patron, however, consults only one subject heading. If that term has no matches, he or she either assumes the library had no materials on the subject or simply gives up.

Steinberg and Metz (cited in Markey, 1984) found that only 28% of the users surveyed were aware that subject searching required them to match an LCSH heading. Whilst Bates (1986) examined this in practice by comparing the degree of match between users' query terminology and LCSH, and found that only just over 20% initially matched or partially matched the controlled LCSH vocabulary. Moreover, there is also the need for users to distinguish as to whether this 'mismatch' is due to 'collection failure' or 'search failure' yet this is not made explicit in many OPAC systems. Users may be unsure as to whether their query formulation failed to match the indexing language of the system (meaning that the library may in fact hold books on the particular topic, only represented by differing terminology); or that the library holds nothing at all for the given topic. The situation often remains ambiguous.

With their card-catalogue, specification-based and exact-match legacy, current commercial OPACs have done little to ease this task. As Markey asserted in 1985, but as pertinent today:

> In our exisiting online catalogues, patrons are saddled with the job of finding the right subject heading to represent their topic of interest.
>
> Markey (1985, p.40)

Despite devices such as search term truncation, acceptance of wild characters, proximity searching and hypertext browsing of subject terms following an initial search, commercial OPACs still require a user to specify and match a subject search. In an evaluation (Borgman, 1986) and re-evaluation (Borgman, 1996), Borgman (1996, p.493) contended that, even with ten year's progress, online catalogues "are *still* hard to use" due to their lack of consideration for the previously discussed information-seeking behaviour research and the legacy of their card catalogue origins:

Most of the improvements are in surface features rather than the core functionality. We see little evidence that our research on searching behavior studies has influenced online catalogue design....Query-based systems were designed from expert librarian searchers who have a rich conceptual framework for information retrieval; their expertise lies in translating questions into queries on behalf of end users...Further improvements in the user interfaces to query-based retrieval systems are likely to have minimal effect on searchers' abilities to answer questions. Instead we need to incorporate more knowledge of searching behavior into the design of these systems.

Borgman (1996, p.501).

Although today, almost another ten years hence, OPAC interfaces have still changed very little. There is little to help a user clarify a vague information need and engage in iterative and exploratory searching behaviour. The OPAC still conforms to Cutter's card catalogue 'objects', divergent from the exploratory nature of subject searching. Yet can the OPAC be moved beyond a finding tool to one of discovery?

3.5 The role of the OPAC: finding or exploratory tool?

Despite the previously discussed research indicating that many users approach information seeking and information retrieval systems with uncertainty, current commercial OPAC interfaces are still oriented towards those users who know exactly what they are seeking and know how to specify this, not the exploratory formulations that characterise subject searching. Such interfaces overwhelmingly conform to the specification led 'form fill-in' model, with little innovation beyond this. Users are typically confronted with a series of indexes (e.g. author, title, title keyword, subject, subject keyword, journal title keyword etc.), based upon fields in the bibliographic record. Users must first select the relevant index, usually from a drop-down list box, and then enter their search terms, or simply enter their search terms in the relevant search index box displayed on the form. Some OPAC interfaces may have a 'General Keyword' index, which searches all indexes simultaneously, but the query formulation is still specification-based. Moreover, although OPACs may incorporate selection via hypertext links within a linear alphabetical display of subject or author headings, for example, or links within a bibliographic record to retrieve related items, this is subsequent to the initial query formulation. At the search outset, there is little in the way of the exploratory features suggested by Marchionini (1995) such as conceptual/organised displays, direct

manipulation and query by recognition and selection, which could aid a user who does not have a specific author or title in mind. As Hildreth urges, we need to:

> break out of the query-oriented, Boolean mind-set, we need to turn the conventional query-first-then-browse paradigm upside down. Searching by exploration, recognition, and discovery in a well structured bibliographic space should be the *primary* (my italics) search interface provided to information seekers...
>
> Hildreth (1995)

Despite this statement, now ten years old, commercial OPAC interfaces are still designed according to the specification approach and the 'finding' of known-items, favouring the user with clearer and more precise information needs than his or her subject-searching counterparts. There is therefore a seeming disparity between the exploratory nature of subject searching and the medium through which this has to be communicated, an OPAC interface based on specification. As Hancock (1987) concluded from her research into users' subject searching behaviour:

> The traditional library and information retrieval system, as we have seen, reduces subject searching to a linear single dimensional sequential process. Online catalogues by contrast should stimulate a more dynamic, multidimensional, truly interactive approach to subject searching and promote a spirit of discovery.
>
> Hancock (1987, p.320)

The dominant 'finding', rather than discovery, model of current commercial OPACs has concurrently restricted the role of classification. No longer is classification utilised for information retrieval, as in the former classified catalogues, but is solely a 'mark and park' tool, a 'locating device' for the library shelves. Despite the classificatory notation appearing on bibliographic records, the exploratory power of its underlying hierarchical structure remains untapped. As will be discussed in the following sections, when employed and manipulated at the system interface, such structures are a possible means of improving subject access and moving the OPAC beyond a mere finding tool to one of exploration and discovery.

3.6 The Untapped Value of Classification at the OPAC Interface

As will be demonstrated in subsequent sections, despite growing recognition of the value of classificatory structures for subject searching on the Internet, commercial OPACs, many of which are now suitably within a hypertext Web environment, surprisingly seem set to remain in 'splendid isolation'. As early as 1964, Swanson (1964) was envisaging an 'automated catalog' with a subject search that used classificatory structures and later as the 'first generation' (to use Hildreth's typology) of OPACs emerged other proponents followed, often demonstrated by the implementation of experimental OPACs, utilising the classificatory structures and/or exploring the potential of the Dewey Decimal Classification Scheme (Markey, 1989) and the Library of Congress Classification Scheme (Cochrane, 1982; Chan, 1986). Moreover, Hildreth (1995) includes provision for 'classification-based searches' within his browsing/exploratory paradigm. More recently, although still assuming a query-based system for subject searching, the IFLA guidelines on OPAC displays (Yee, 2003) do recommend that a user be shown the cross-disciplinary classification categories for a specified search term and be able to 'right click' on the mouse to view the hierarchical context of any classification number.

In commercial systems, however, classification schemes are used for the physical purposes of *directing* users to items on shelves, with access to the hierarchical knowledge structures behind classification schemes remaining the sole preserve of the librarian.

> ...classification does not play any significant role in many OPACs and that the lessons learned over many years from classified catalogues have not been incorporated in many online systems. It is not that the subject approach used in the classified catalogue is perfect - far from it, there are numerous difficulties - but it is patently clear that the use of other subject approaches in OPACs has not led to any improvement. Indeed in some cases subject retrieval is far worse.
>
> Hunter (2000)

Considering the wealth of intellectual effort involved in assigning classificatory notations, as Cochrane (1982) argues, this seems "a horrible waste", particularly, as previously discussed, when users have been found to experience many problems when bringing vague and evolving information needs to query-driven subject search options

(Markey, 1984). Based on the findings of studies by numerous researchers, Cochrane and Johnson (1996, p.95) argue:

> In our opinion, the technologies at our disposal at the time were not sufficiently advanced to provide the functionalities required. What is frustrating now is to find that system designers today, with even better technologies, have learned little or nothing from those early OPAC user studies, from the analytical papers on preparing classification and thesauri for use online..., from the early attempts to mount thesauri and classification systems into retrieval/search systems...

Little has changed since Cochrane and Johnson made this assertion. In current OPACs, the system knowledge structures are 'hidden' from the user, whereas a visual display of classificatory structures could resolve some of the problems users experience when subject searching, blindly attempting to match their own search vocabulary to that of the system.

> Information retrieval (IR) is only effective when the people involved share common knowledge structures. The match of concepts among user, intermediary, and system then has some chance of success.
>
> O'Brien (1994, p.219)

If the user had visual access to the knowledge structures used by the library system then perhaps subject searching would be more fruitful, and browsing encouraged, due to the accommodation of search by selection as well as specification. As a result, this might help users formulate a search; easily narrow or broaden a search, if too many or too few items respectively are retrieved; show the context and terminology of a subject discipline, displaying related terms and disambiguating homophones; and facilitate exploration, browsing and discovery (Vickery, 1990; Cochrane and Johnson, 1996; Koch, 1997; Svenonius, 1983; Vizine-Goetz, 1997, 2002).

Making the knowledge structures of classification schemes visible at the OPAC interface may also serve to have positive impact on students' learning experience and even promote creativity. Jack Mills is an ardent advocate of the educational value of classification:

> ...classificatory structures assist seekers of information to realize the connectedness of concepts in a store of information, this must have implication for the education of those seekers... The value of a classification in formulating a search prescription is the result of its systematic presentation of the connectedness of concepts. It presents a clear picture not only of the concepts involved but also of their generic contexts and their syntactic relation.
>
> Mills (1997, p.9).

Mills cites the work of Lilley's 1954 study in which students were asked to specify subject headings for a particular search area. The results "in many cases demonstrated a serious absence in their education of any sense of an overall structure of knowledge or of any sort of connectedness between different fields which might give them some sort of perspective when searching for information" (p.10). The informative context of classificatory structures for searching and browsing in a library OPAC may go some way towards improving a situation which Mills believes has changed little since completion of Lilley's study. Furthermore, other researchers (Bawden, 1986; Davies, 1989; Kwasnik, 1999) have suggested that searching using classification can encourage and reveal serendipitous links, discoveries and the creation of new knowledge.

The possible benefits - assisting subject search formulation, browsing, providing an overview of knowledge, indicating conceptual relationships, and encouraging creativity and discovery – of displaying the conceptual structure of the classification schemes which organise physical library collections is thus unrealised in the electronic domain of commercial OPACs.

> Structured browsing is, in other words, as important as searching, especially when the user is not sure what he or she is looking for. OPAC systems are not designed to provide this approach. They can meet part of the need, namely for searches limited by resource type. But the hierarchical browsable directory that it now a standard feature of most Internet search engines and eLib subject gateways is, I believe, not possible with the current generation of OPACs.
>
> Hanson (1998)

However, as Hanson asserts, classificatory principles are being exploited in the Internet domain.

3.7 Classification at the Internet and Online Interface

It is perhaps ironic that, as we have seen, classification owes its adoption, development and unfortunate relegation to how materials should be organised and retrieved within a library environment, yet it is now being re-discovered and re-cast by the Internet community, a computer culture which had initially promoted the exact-match query-

oriented paradigm. The advent of graphical user interfaces, hypertext and the World Wide Web provided the opportunity of a browsable interface.

3.7.1 'Home-Grown' Subject Directories and Standardised Subject Gateways

Despite a specification-based keyword search still being the most predominant and, arguably, most popular method of searching the Internet (viz. Google), some search engines (e.g. Google, Alta Vista, Yahoo) also adopt a classification-based approach, presenting users with hierarchical categories from which to choose. Although such subject directories, covering all subject areas, have been sometimes criticised for their lack of logical structure, inconsistency in the degree of specificity, insufficient context and indication of related terms (Vizine-Goetz, 1998), when compared to traditional classification schemes, they do at least provide an alternative means of more exploratory searching.

An, arguably, more rigorous and librarian-led venture is the popular Resource Discovery Network (RDN, 2004) of eight subject 'HUBs'/Gateways (e.g. Artifact for Art and Design, SOSIG for social sciences, HUMBUL for humanities etc.). All promote subject browsing via classification-based categories, underpinned by the consistent meta-data standards of Dublin Core cataloguing elements and often standardised subject headings, classification schemes and thesauri for retrieving and, in some cases, organising content. For example, the HUMBUL gateway has investigated use of the Library of Congress Classification Scheme (Fraser and Edwards, 2001); PSIgate, the Physics and Astronomy Classification Scheme of the American Institute of Physics (PSigate, 2004); and EEVL, the Ei Thesaurus, Mathematics Subject Classification and ACM Computing Classification System (EEVL, 2004). Similarly, Artifact utilises the Getty Art and Architecture Thesaurus for subject keywords (whilst using its own Artifact list for subject headings/categories) and also assigns Dewey codes to records (Artifact, 2004); and BIOME (2004), the health, life sciences and medical gateway has organised its content using a combination of selected 'top-level' National Library of Medicine headings, Library of Congress Subject Headings and Dewey headings, depending on the subject area.

Both subject directories and gateways are therefore exploiting the value of incorporating a browsable, category led structure in addition to keyword searching. Internet search engines rely upon their own 'home-grown' schemes, whilst standardised classification, thesauri and subject indexing schemes are employed within the RDN to varying degrees, either by combining them with a HUB's own 'home-grown' scheme, adopting several different standardised schemes for organising different subject-specific areas within a broader subject gateway; or using such standardised schemes as a 'basis'. As demonstrated above, the majority of these RDN gateways use subject-specific schemes and those employing the universal schemes of the Dewey Decimal Classification Scheme or Library of Congress Subject Headings only employ the relevant part. The following section investigates online applications of standard universal classification schemes for organising resources from numerous, not discrete, subject disciplines.

3.7.2 Enumerated/Hierarchical Classifications

The emergence of computer technology and bibliographic classification schemes becoming available in machine-readable form saw early experiments into the use of classification for online information retrieval. The first of these was an application of the Universal Decimal Classification scheme (UDC), as described in Chapter 2, named the AUDACIOUS system and presented, initially, at the Conference of Mechanized Information Storage, Retrieval, and Dissemination in 1967 (Freeman and Atherton, 1967) and then, more fully, a year later (Freeman and Atherton, 1967; Cochrane, 1982). This system enabled the user to formulate a query using one or more keywords, or a UDC notation if known, to retrieve a list of UDC schedule captions, containing the desired keyword. The user could then select a caption and 'scan' its hierarchical structure in context, using the relevant segment of the UDC schedule. The pioneering AUDACIOUS system, therefore, initially operated as a conventional specification-based information retrieval system, albeit with a controlled vocabulary, but then allowed a user to view related terms (broader, narrower) in their hierarchical context to provide added confirmation that the search was as desired or additional/alternative terms which could be

saved for subsequent searches; searches could not be reformulated in the hierarchical view. Following preliminary testing, the authors conceded that the system would need "extensive modification" to be accepted in a real environment (Freeman and Atherton, 1967, p.347). However, it was believed that the system had demonstrated its primary objective: that it is technically feasible to use a classification scheme for online information retrieval and that this had favourable ramifications for the application of other bibliographic classifications. However, it was not until the emergence of online public access catalogues in the early 1980s that this began to be realised.

In 1984, following evidence of the prevalence of subject searching and the problems users were encountering, a project was commissioned by the Online Computer Library Centre (OCLC), owners of Forest Press and publishers of Dewey, Office of Research to investigate the use of the Dewey Decimal Classification Scheme as an online search tool (Markey, 1985, 1987, 1989). Technology had advanced since the implementation of AUDCIOUS (Freeman and Atherton, 1967) and the classification scheme could be used, not just for entry vocabulary and hierarchical context display, but for broadening or narrowing *within* a search interaction. Query formulation, however, was still wholly specification based, with users entering a subject search term that is then best matched to terms within the DDC captions, relative index, notes and the first heading in the bibliographic record. If the user's search term matched more than one caption, then all 'perspectives' were displayed from which the user could choose the most relevant. Whether the search matched one or many captions, the user could manipulate the results by displaying broader/narrower captions, related information from the Schedules or Relative Index, or the retrieved items. In system tests, despite the extensive knowledge base and cross-referencing, users often experienced problems in matching their own vocabulary to that of the Dewey Online Catalogue (DOC) system, often being directed to too broad a caption primarily due to system's complex 'grouping' procedure. In the event of multiple Dewey caption to search term matches within the same Main Class, this procedure directed users to only the shortest notation (i.e. broadest class), as opposed to the caption with the largest number of items, which was often confusing or appeared irrelevant to users. Other issues included instances when users misunderstood the

terminology of the initial display of 'perspectives', without the surrounding hierarchy; and in multifaceted searches, users felt the system could represent a single facet of their query effectively but wished to limit their results by subsequent facets such as year of publication, geographic area, language etc.. All of the above issues, although particularly the latter, have implications for the current PhD research and will be considered in subsequent chapters. The DOC system, however, received encouraging feedback from users, which although was not designed to help users formulate their initial search, did help refine the search, suggest related terms and provide a logical and informative context for searching. More recent research and applications have utilised enumerative classification schemes from the outset, to help users 'formulate' a search.

As stated previously, the arrival of the Internet, the graphical user interface and hypertext enabled traditional classifications to assume a new role: the organisation of electronic resources and direct online manipulation of the scheme's structure for searching and browsing such resources. Another advantage of classification in this new online role, as alluded to at the end of Chapter 2, is that it is freed from the physical constraint of a single relative location and the scheme can be more easily manipulated to provide alternative arrangements. There are several applications of universal bibliographic classification schemes, primarily the DDC, on the Internet; however, these applications are perhaps not exploiting the electronic domain to its full advantage. Since the DDC is the most predominately used scheme and central to the current research, it will be the focus here.

Examples can be found of DDC summaries being transplanted, essentially in their paper-based form (although some use slightly modified captions) together with the decimal notation, in subject gateways such as:

- *BUBL Link*
  < http://bubl.ac.uk/link/>
- *Canadian Information by Subject* at the National Library of Canada
  <http://www.nlc-bnc.ca/caninfo/ecaninfo.htm>
- *CyberDewey* by David Mundie
  < http://www.anthus.com/CyberDewey/CyberDewey.html>
- *Net Sites by the Numbers* at Tempe Public Library, Arizona
  <http://www.tempe.gov/library/netsites/>
- *Webray* at Morton Grove Public Library, Illinois
  <http://www.webrary.org/ref/weblinksmenu.html>
- *The UK Web Library – Wolverhampton Web Library (WWLib)* by Peter Burden
  <http://www.scit.wlv.ac.uk/wwlib/browse.html>

In these gateways, Dewey is utilised in a menu-driven interface, with the user selecting a single Dewey class and browsing down/up each hierarchy via a series of screens until the desired Internet resources have been located. Non-Dewey caption direct searches are also usually available via a form-based interface. With 12,000 selected Internet resources, BUBL Link is perhaps the best known and most widely used of the above subject gateways in the UK. Currently, it assigns one Dewey code per resource. In a previous paper (Tinker et al, 1999), it was suggested that multi-faceted resources could be assigned multiple Dewey codes to enable access from different disciplinary viewpoints, with the proviso that there was a policy to determine how exhaustive this 'classification as indexing' would be, or a weighting scheme to indicate the relative importance (extent of coverage) of each descriptor. Beghtol concurs with this practice of classification as pre-coordinate indexing:

> Bibliographic classification systems express multidisciplinary topics easily if one assumes a classified catalogue (manual or electronic) with multiple notation access points for each document
>
> Beghtol (1998b, p. 4).

For example, *Iter: Gateway to the Middle Ages and Renaissance* (Iter Inc., 2005), cited by Beghtol, assigns multiple Dewey codes to online journal articles and reviews to improve access to interdisciplinary materials. Davies (1989) claims that the ability to search across disciplines, or 'horizontally', may also stimulate creativity and quicken the resolution of research problems by connecting researchers in different disciplines

working on analogous research questions, previously unrealised due to differences in terminology.

Thus, in an electronic context where there is no requirement to locate an item in one single place, multiple Dewey notations and their accompanying captions can be assigned to enable a user to access a subject from a variety of angles. As Searing states:

> The classification structure is a theoretical map of human knowledge, but in practical terms, it is a map of the stacks, guiding readers to places where works on similar topics can be found in physical proximity. In the electronic environment, texts are freed from the limitations of physicality, so in theory any number of classes can be assigned.
>
> Searing (1996, p.320)

In addition to the possible assignment of multiple Dewey codes between and even within Main Classes, a single Dewey code itself contains multiple subject components/facets, as will be demonstrated in Chapter 4 and subsequent chapters, but in the physical domain access is only provided via the first. For application in the electronic domain, such classificatory notations can be deconstructed to reveal the embedded facets, providing access and collecting together according to alternative viewpoints. Although more difficult with more rigid hierarchical classifications, such flexibility is inherent within faceted classifications. The application of these for providing a physical arrangement was considered in the previous Chapter 2; the following section examines their application within an online domain.

3.7.3 Faceted Classifications

As discussed in Chapter 2, faceted classification provides a more flexible approach to knowledge organisation than its preceding 'top-down' hierarchical schemes. Whilst more easy to manipulate and accommodating complex subjects more readily, the entrenchment of hierarchical schemes appears to have curtailed its adoption in the physical domain of shelf organisation. However, there appears to be a growing interest in the online environment.

An early online application of faceted classification was in the Data Resource Directory (DRD) of the United States Energy Information Administration (Travis, 1982). The classification was a special scheme (i.e. not Universal), developed by the Energy Information Administration and called the Classification Scheme for Energy Data, comprising eight distinct facets, three of which contained another eight sub-facets in total. Despite constraining the facets to a fixed citation order "to minimize ambiguity" (Travis, 1982, p, 272), an order which is not essential in an online domain, the system did display an awareness of the flexibility of the online environment:

> In an online system, one need not be so rigid about crosslisting concepts, provided that the context does not change the meaning of the term enough to require a new heading. For example, "synthetic natural gas (coal)" is listed under both "synthetic natural gas" and "coal products".
>
> Travis (1982, p.274)

Thus, unlike a classification system applied in a physical context, restricted by the need for a single relative location, the DRD system exploited the potential of the online domain, by placing terms under multiple headings, as appropriate, and, in turn, allowing multiple entry points when searching. Although the search interface itself was predominately specification based, it did attempt to translate users' search terms to that of the controlled vocabulary, and provide an A-Z (VOCAB) and conceptual listing (BROWSE) of the classificatory structure, showing which terms were already or not in the current search, by way of plus and minus signs respectively. A user could then elect to add or remove terms from the search. The system employed implicit Boolean logic, with search terms within the same facet being automatically OR-ed and those in different facets AND-ed, enabling cross-disciplinary searching. For its time, the system appears relatively visionary, but more widespread adoption was perhaps affected by the attitude to classification in the United States, as mentioned in Chapter 2. Interestingly, the authors concede that, although the system was based upon faceted classification, they needed to disguise it as "a conventional alphabetical thesaurus":

> Although the database and indexing would support classified displays, we elected not to build them. This decision was influenced by the fact that classified catalogs are unknown in the United States, not only to the public, but to most information professionals. There was a negative reaction among some EIA staff even to the notation in the scheme. The user education problem in gaining acceptance for an online classified catalog in any conventional sense of the term would have been formidable. Our approach therefore suppressed the notation from the interface.
>
> Travis (1982, p.273)

Although a notation is not necessarily warranted in an online interface, since it serves to indicate a physical shelf location, the above quote highlights the extreme opposition or simply ignorance of classification, let alone faceted classification, for information retrieval purposes in the influential United States at this time. However, twenty years later, with a handful of research prototypes along the way (e.g. Allen, 1996), the United States web community are beginning to discover the value and principles of faceted classification; but this time making it explicit and making it their own.

This recent interest in moving Internet subject browsing from the linear structure of the hierarchical classification that restricts movement to upwards or downwards within a single hierarchical structure at one time, as described in the previous section 3.7.2, to one which provides multidimensional access across several hierarchical structures simultaneously is now being demonstrated in various web applications (e.g. Endeca, 2004[8]; Facet Map, 2003) and is capturing the attention of the knowledge management industry (Knowledge Management Connection, 2004; La Barre, 2004). Interestingly, in developing their self-called 'Guided Navigation' technology, Endeca - a name derived from the German verb entdecken: to discover (Endeca, 2004) - were clearly aware of research into the problems of subject access and the limitations of purely specification-based interfaces, as described previously:

> Guided Navigation eclipses conventional approaches that rely on the slow and often confusing method of having to ask the right question, and then waiting for what may or may not be a useful reply. This interaction model, which we refer to as 'query and response', puts a burden on users to be familiar with the data set so they have an idea what to look for, and how to describe the item and structure the query in a way the system will recognize. Because of the difficulty inherently in this approach, "query and response" systems, which include keyword search parametric search, SQL and others, typically return either a useless long list of results or the frustrating "No Results Found."
>
> Endeca (2004)

These problems sound incredibly familiar; and it is faceted classification that is being utilised to address such problems. Although not employing standard bibliographic faceted classification schemes, such as Bliss or Colon, these Internet applications are

---

[8] An interesting application of Endeca 'Guided Navigation' technology is by Barnes and Noble Bookstore's Bookbrowser < http://www.barnesandnoble.com/>.

creating a browsable selection-based interface that adopts their principles. Such applications present individual facets of a subject area and then assign an item to one or more facets, as appropriate, corresponding to the various subject dimensions of that item. Thus, an item is classified and retrievable according to any of its facets in any citation order. Moreover, a user can also combine and view the results of multiple facet hierarchies (e.g. Subject Discipline, Geographic Location, Time Period etc.), browsing up and down each to broaden or narrow a search. Consequently, they can provide the kind of "n-dimensional" access that Ranganthan (1965) could only envisage, not constrained by the physical, one-dimensional restrictions of the notational plane, where a facet citation order is mandatory. This multidimensional navigation brings to mind an analogy from a classic children's book, where, similarly, conventions do not apply:

> 'If you're tired then we'd better take the lift', said Mr Wonka. It's over here. Come on! In we go!...This isn't just an ordinary up-and-down lift!' announced Mr Wonka proudly. 'This lift can go sideways and longways and slantways and any other way you can think of! It can visit any single room in the whole factory, no matter where it is!'
>
> Dahl (1964, p.130)

Thus, the Internet community, akin to their earlier adoption of hierarchical classification in subject gateways such as Yahoo and Google, are similarly now beginning to discover the advantages of a faceted approach, with their own 'home-grown' schemes. Although clearly an encouraging trend, La Barre (2004) believes that there is still much confusion amongst the Internet and Knowledge Management communities as to a basic definition of faceted classification and urges the library, classification and knowledge organisation communities to ensure their involvement:

> This is not a time to stand back and watch disinterestedly, it is a time to plunge in headlong. It is not enough that facets are a hot topic on the discussion lists of IAs and KM specialists. Hot topics can fade as quickly as the systems they spawn begin to fail. Systems built with incomplete understandings of the theory they seek to actualize, can serve as fodder in an all too common argument that failures belong not to the implementation themselves, but the underlying theory.
>
> La Barre (2004)

Therefore, according to La Barre, the acceptance and success of such applications, and ultimately faceted classification *per se*, will be influenced by the degree of rigour and theoretical understanding employed within the systems by their creators. The

library/research community is involved, perhaps to a more limited degree, in investigating faceted classification at the online interface, but favouring standard bibliographic schemes.

Researchers working within the Knowledge Organisation community are involved in enhancing existing faceted classification schemes for the purpose of online searching, creating systems which apply standard faceted classification schemes for searching and browsing at the online interface, and evaluating users' interactions with such systems. For example, a project based at University College London (Broughton, 2001b; Broughton, 2002; Broughton and Lane, 2004; School of Library, Archive and Information Studies, UCL, 2003) has recently investigated the application of Facet Analytical Theory - developed by the Classification Research Group (see Chapter 2, section 2.6) - in the online searching/browsing of Internet resources. The project seeks to develop a new faceted classification, based on an amalgamation of the "best features" of the Bliss Bibliographic Classification 2, the Universal Decimal Classification and the Broad System of Ordering. This new faceted classification scheme, named FAT-HUM, is a special classification, focusing solely on the humanities. It is planned to be applied in the convergence of two major subject gateways (both currently using hierarchical/enumerative classification schemes), namely the previously described HUMBUL (2005) and the Arts and Humanities Data Service (2005) to create a single Humanities portal, although this has not occurred to date (Broughton and Lane, 2004). The system is based on the following premise, highlighted previously as a valuable opportunity for faceted classification within an online domain, and will be tested for its effect in promoting cross-disciplinary searching:

> What is peculiar to the digital context is the search facility and lack of need to adopt a linear arrangement. This can allow us to capitalise on the complexity of the document description and the potential multiplicity of access points.
>
> Broughton (2001b, pp.89-90)

Another Internet-based application that utilises a standard controlled vocabulary is the Flamenco (FLexible information Access using MEtadata in Novel COmbinations) search interface project, led by Marti Hearst at the University of California, Berkley (UC Berkley, n.d.; Yee et al, 2003). The research team refer to their approach as "Faceted

Metadata", utilising, not a faceted classification scheme as such, but subject terms from the Art and Architecture Thesaurus, developed by the Getty Research Institute. There are two versions of this image system: one covering fine art and the other architecture.

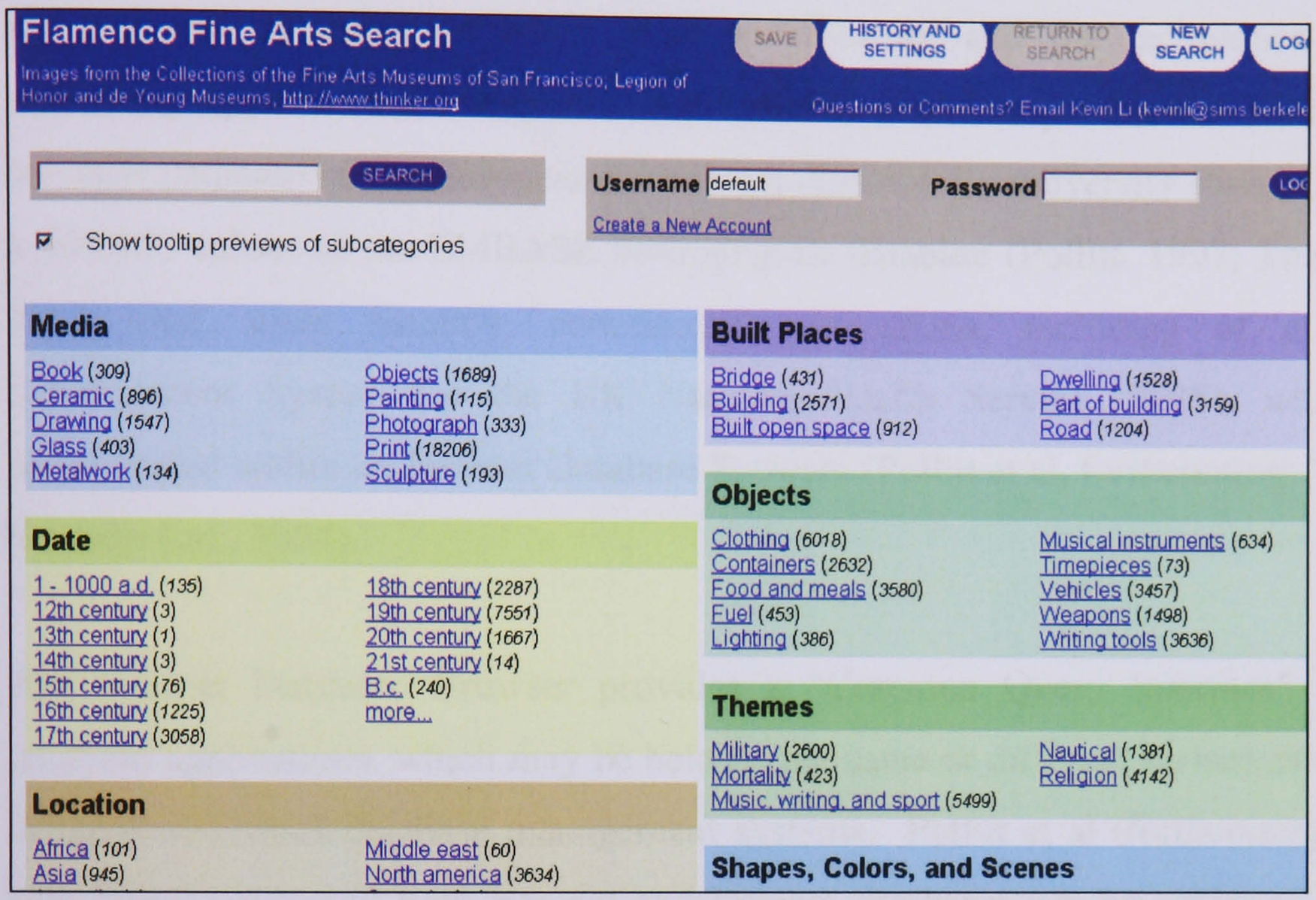


Figure 3.2: Flamenco Fine Arts Search Interface (UC Berkley, n.d.)

As can be seen from Figure 3.2, the user is presented with various facets, together with the corresponding number of images. The user can begin a search by entering a search term or selecting a category within a facet (for example, Media>Painting = 115 images). Other facets can then be added to this search and their hierarchies browsed, to constrain the search in an implicit Boolean AND-ing of the facets (e.g. Location>Europe, reduces the set from 115 to 72 images; then, Date>18th Century, reduces the set from 72 to just 8 images). A search can also be broadened by removing a facet. This ability to search by selection and progressively refine or broaden a search by multiple category dimensions was well-received in user-testing (Yee et al, 2003), as will be described further in the forthcoming usability study in Chapter 9. Indeed, of the systems reviewed in this section so far, the approach of the Flamenco interface is the one that most closely resembles our own View-Based Searching interface (Pollitt, 1997), the interface which will later exploit the research undertaken in this current PhD, as will be detailed in subsequent chapters.

Underpinned by a relational database model, View-based Searching utilises knowledge structures in navigable views, reflecting facets of the objects in databases with implicit Boolean searching and mutually constraining views. View-based Searching interaction is characterised by selection rather than specification and the techniques have been previously applied in a number of applications: a directory of hotel information, the business database of the European Parliament (EPOQUE), university student records and a 600,000 subset of the EMBASE bibliographic database (Pollitt, 1997; Treglown et al, 1997) and, more recently, commercial applications, including of a Workforce Development System for the UK National Health Service (NHS), which can be incorporated within an Internet Database Browser (Pollitt et al, forthcoming; View-Based Systems Ltd., 2004).

The Internet Database Browser provides a "Common Query Interface" to multiple database applications, which may be held on the same or different servers and even have different relational database management systems. Pollitt et al (forthcoming) state that this generic access to both Internet and Intranet databases can be achieved by defining metadata for applications, objects and views (facets). Such database applications currently include a 'local' NHS Workforce Development System, as stated above, and also a View-based searching interface to the 'worldwide' Internet Open Directory Project, a classified arrangement of over four million Internet resources, hosted by Netscape and underpinning many of the major search engine directories, e.g. Google, Lycos etc. (Netscape, 2002; Netscape, 2004).

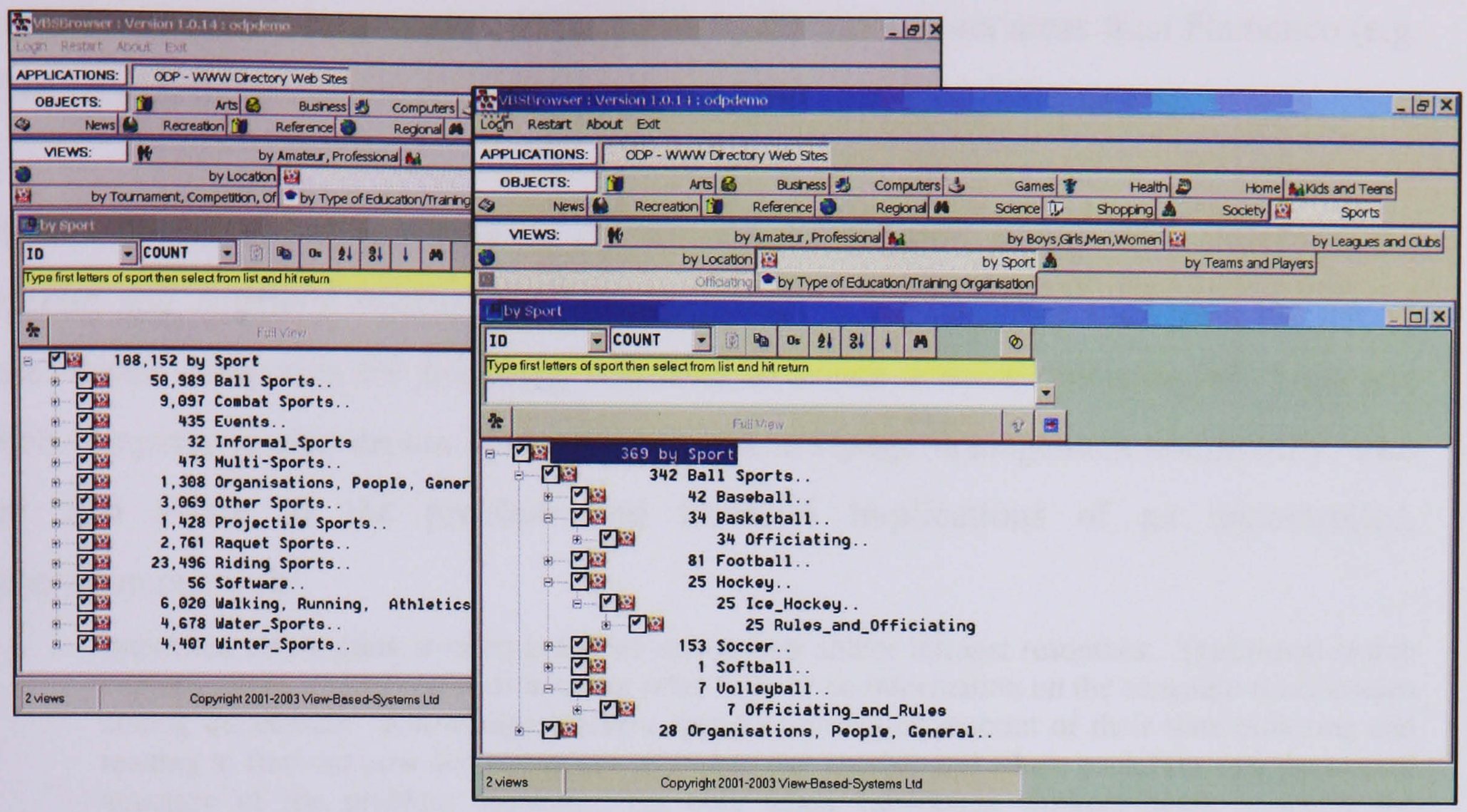


Figure 3.3: View-based Searching Interface to the Open Directory Project

Figure 3.3 shows the user initially selecting the 'By Sport' view (facet), then focusing on 'Ball Sports' combined with the 'Officiating' view; this narrows the search from 108,152 websites to 369, which could then be refined further by selecting the particular type of ball sport and, ultimately, the resultant website displayed within a web browser. Thus, when browsing, such navigation is enabled by a 'point and click' interaction with the user moving up and/or down each hierarchical view. The views are linked, mutually constrained by an implicit Boolean 'AND' query, and allow the user to select persistently the degree of specificity for each particular facet, simultaneously observing the effects across all facets.

Despite the approach to facet-based searching in both View-based systems and Flamenco being similar at the interface (see Chapter 8 for View-based interface utilised in this current research), there are differences. View-based searching systems currently have a different platform to the Flamenco system, in that Flamenco is web-based, whilst View-based searching is a windows-based application, although, as we have seen, it can provide a searchable interface to Internet data. A web-based version, however, has recently been investigated and several different prototypes have been produced with extremely promising results. Another difference is that View-based searching has

typically dealt with larger collections and more diverse subject areas than Flamenco (e.g. VBS EMBASE containing 600,000 records and Flamenco, 60,000 images).

Parallel to the above applications of faceted classification at the Web interface, the Internet and Artificial Intelligence (AI) communities are currently exploring the use of facet-based categories for encoding metadata to create a more "meaningful" Semantic Web. Impetus is also driven by the corporate knowledge management community, who are also aware of the problem and financial implications of an uncontrolled, heterogeneous web:

> Important information is often scattered across web and/or intranet resources. Traditional search engines return ranked retrieval lists that offer little or no information on the semantic relationships among documents. Knowledge workers spend a substantial amount of their time browsing and reading to find out how documents are related to one another and where each falls into the overall structure of the problem domain. Yet only when knowledge workers begin to locate the similarities and differences among pieces of information do they move into an essential part of their work: building relationships to create new knowledge.
>
> Davies et al (2003, p.3)

These activities of knowledge workers are resonant of the practice of classification in their ascertainment of document relationships, "similarities" and "differences" and an "overall structure" of a domain. The value of a classificatory approach is clearly of relevance, although the different information and knowledge communities adopted alternative implementations. Whilst librarians sought to rationalise the web for users by creating explicit, browsable subject gateways and directories of selected resources (as previously discussed), the AI and Knowledge Management Communities focused on the underlying infrastructure and how additional technological "layers" of meaning (Koivunen and Miller, 2001) might be encoded into web documents. The focus is to develop and standardise metadata (structured data about data) for web resources, almost akin to bibliographic records of library items in a library OPAC, enabling relationships and machine communication amongst websites that was previously limited (i.e. hypertext links), as Koivunen and Miller explain:

> While a knowledgeable human may realize easily that one resource is conceptually an invoice and another one is a novel or a research paper this information is often unavailable for a machine. Similarly a user can guess what kinds of relationships the resource has by reading the text around the link, but it is hard for the machine to make these same guesses. More informative relationships would be, for instance, "depends on", "is a version of", "has subject", "authors".
>
> Koivunen and Miller (2001)

This vision of the Semantic Web seeks "to *evolve* the current Web to one where information and services are understandable and usable by computers as well as humans - to create a 'Web for machines'" (Goble, 2003, p.551). According to the World Wide Web and Semantic Web pioneer Tim Berners-Lee:

> The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation...Human language thrives when using the same term to mean somewhat different things, but automation does not.
>
> Berners-Lee et al (2001)

It appears, then, that the rigour of classification is being discovered and re-branded as a valuable tool for disambiguating and creating, in bibliographic terms, universal vocabulary control of the Web. This will move the Web from a series of resources and hyperlinks to one in which relationships are recognised, defined and encoded to enable more advanced machine communication. Semantic Web researchers claim that this will facilitate organisation of the huge diversity of documents on the web, improve the accuracy of search engines, facilitate cross-database searching, sharing of data and allow agents to perform tasks for the user, based upon this classified metadata. In order to enable the Semantic Web and its metadata, however, several tools have been and are still undergoing development, coordinated by the World Wide Web Consortium (W3C) (World Wide Web Consortium, 2006a).

These tools comprise several independent technological layers, varying in function, sophistication and similarity to library classification. At the foundation level is eXtensible Markup Language (XML). Developed in the 1990s and accepted as a standard by W3C in 1998, XML was the first move towards metatdata in that, in contrast to HTML which determines how information will look on a webpage, it describes and encodes the type of document (World Wide Web Consortium, 2006b; Passin, 2005). The more advanced XML Schema later followed, defining a standard set of elements and attributes for the content and structure of a particular document type. It was the subsequent Resource Description Framework (RDF) and RDF Schema, however, which began to establish semantic relationships about the subject content of documents more

akin to those of library cataloguing and classification. Again, RDF is a mark up language, yet it categorises the content of documents in a logical statement of subject, predicate and object. RDF Schema formalises this RDF vocabulary into classes - the web resource itself constituting a class and uniquely indicated by a Universal Resource Identifier (URI) - and related properties (including the aforementioned subject, predicate, object 'triples') that can define hierarchical structures of sub-class and sub-property (World Wide Web Consortium, 2004a; Wilson, 2004). Whilst the World Wide Web Consortium were developing RDF and RDF Schema, another group of researchers (e.g. Pepper, 2002) focused upon creating related topic maps, utilising XML (XML topic maps, XTM) to allow document interchange. This work culminated in an International Standards Organisation (ISO) standard 13250 (International Standards Organisation, 1999). The focus of the standard is to provide:

> ...a standardized notation of the interchangeably representing information about the structure of information resources used to define topics, and the relationships between topics. A set of one of more interrelated documents that employs the notation defined by the International Standard is called a *Topic Map.*
>
> International Standards Organisation (1999, p.iii)

There are three main components in a topic map: topics, occurrences and aesociations (Pepper, 2002; TopicMaps.org, 2001). 'Topic' is similar to a subject in library terms and can be classified into more specific 'topic types'; each topic is usually given a 'topic name'. In library classification, such 'topic names' are generally the equivalent of a controlled vocabulary of classificatory concepts (captions of preferred terms) with synonym cross-references (e.g. the Relative Index in Dewey); whereas, topic maps have similar vocabulary control but allow the flexibility of different topic names for different contexts (scopes). Thus, since the creator of the topic map can assign a number of different topic names to the same topic, this could enable an end user of a website (encoded by a topic map) to set up a profile of their preferred terminology (topic names) (Garshol, 2004). Moreover, Garshol also explains how different topics can also be assigned the same topic name, in contrast to the "disambiguation" required of library classification:

> Many classification systems avoid duplicate names by including some disambiguation in the name itself. For example, Paris the city and Paris the hero of Greek mythology may have been differentiated as "Paril (France)" and Paris ("Greek myth.)". In topic maps this is not necessary, and the types, occurrences, and associations of the topics will generally distinguish them anyway.
>
> Garshol (2004)

'Occurrences' refer to actual instances of information sources (although a topic map can be created without reference to a document collection), pointing from 'topics' to relevant documents, showing strong parallels with the act of library classification. 'Occurrence Roles' can also be specified, equivalent to a library's document type (e.g. journal article, book, etc.) 'Associations' are much more specific than the primarily hierarchical relationships in the majority of library classification schemes with a large range of possible 'Association Types' and 'Association Roles', which relate concepts (topics) and documents (occurrences) across the topic map. Although inspired by a drive to create a tool equivalent to a printed back-of- book index for the online domain (Pepper, 2002), topic maps demonstrate striking parallels with faceted classification; the ISO standard refers to the topic maps as a multidimensional space with topics connected by associations (relationships) and even the term 'facet' is used and defined as "a property of a set of information objects that can be used to create a view of them" (International Standards Organisation, 1999, p.iii). Pepper (2002) argues topic maps can play a key role in navigating information resources, with metadata being assigned to sources (occurrences) in the form of such facets, enabling "query filters producing restricted subsets of resources, for example those whose language is "Italian" and user level is "secondary school student". The result is akin to how classification collates (associates) documents (occurrences) on the basis of conceptual similarity/subject (topic), according to common properties (facets). Whilst classification manages the multidimensionality of documents by encoding the various components within a pre-coordinate classificatory notation, yet then only explicitly representing the first/primary concept due to the constraints of linear physical access, the online world need not adhere to such constraints. Topic maps can adopt a more flexible post-coordinate approach, allowing a resource to have multiple topics (and therefore access points) and even multiple topic maps for the same information resource (International Standards Organisation, 1999; Pepper, 2002). However, although more specific in their terminology and encoding richer relationships

than a library classification scheme, they focus upon a particular domain of knowledge for a particular community of users, not aiming to represent the entire universe of knowledge, as in a library classification scheme. Despite clear contrasts, topic maps demonstrate a particular resonance with library classification in that they are concerned with the management and retrieval of information resources (objects), a closer relationship than the previously discussed RDF and RDF Schema. Garshol (2002) argues that although the appearance of topic maps and RDF are seemingly similar, RDF is more suited for supplying "fine-grained metadata" whilst topic maps "for making information more findable". Similar to topic maps, a further tool with a key role in the Semantic Web is the ontology. Developed alongside topic maps but by the AI community (coordinated by W3C), ontologies also seek to provide a shared conceptual representation and relationships within a specific knowledge domain. Although conceived as early as the 1980s (yet the philosophical term 'ontology' is thought to originate from ancient Greece), ontologies have generated much interest since recognition of their perceived role in the vision of the Semantic Web (Wilson, 2004). Constituting a higher, more sophisticated 'layer' in the Semantic Web than XML, RDF and RDF Schema, ontologies encode 'Individuals' (e.g. things, objects, documents, occurrences etc.), 'Classes' (topics/subject of the things, objects, documents), 'Attributes' (properties of the objects) and 'Relationships'(how the objects relate to each other) (World Wide Web Consortium, 2004b). These classes, attributes and relationships are created by virtue of the logic-based Web Ontology Language (OWL), which builds upon and advances RDF and RDF-Schema. As in the case of topic maps, relationships within an ontology are more expressive and specific than a library classification scheme, as Koo and Na define:

> In practice, an ontology is expressed as a taxonomy of concepts linked by IS-A, part-whole, and attribute-value relations, sometimes enriched by other kind of relations as well as additional rules or constraints called *axioms*.
>
> Koo and Na (2006, p.183)

Koo and Na explain that the number and type of relationship can vary across different ontologies and domains; for example, they cite the ontology of the Unified Medical Language System (UMLS) based at the USA National Library of Medicine, with its:

> ...54 relations grouped broadly into IS-A and associated_with relation types, the latter being subdivided into physically_related_to, spatially_related_to, functionally_related_to, temporally_related_to, and conceptually_related_to.
>
> Koo and Na (2006, p.187)

However, although creating richer relationships than a library classification, Koo and Na (2006) claim that there has been little documentation of the numerous relation types used across different ontologies, documentation which would be recorded and available in the more controlled and standardised vocabulary of a library classification scheme, being much more established. To encourage such standardisation and associated interoperability, rationalisation and further coordination may be necessary for progression amongst this proliferation of separate initiatives. A recent proposal has been developed for data interoperability between topic maps and RDF/OWL ontologies to unify the two schemas (World Wide Web Consortium, 2006c) and a move towards ontology mapping of separate domain specific schemes to enable universal interoperability. Furthermore, there is growing recognition that the specific vocabulary, relationships and inference possibilities of OWL may not be a panacea for all types of domain. The AI community is looking to draw upon the intellectual effort of library classification schemes and their related collections to bring them into the remit of the Semantic Web (Herman, 2005).

The emergence of ontologies in the 1990s led Soergel, from the Knowledge Organisation community, to question whether they were simply "a reinvention of classification":

> Classification serves many functions and thus claimed by many fields, but the communication among these fields is poor, leading to an approach that is marked by fragmented and costly reinvention...Classification has long been used in library and information systems to provide guidance to the user in clarifying her information need and to structure results for browsing, functions largely ignored by the text retrieval community but now receiving increasing attention in the context of helping users to cope with the vast amount of information on the Web.
>
> Soegel (1999)

However, as we have seen, although there are clear parallels between ontologies, topic maps and library classification in terms of the basic act of classifying (grouping my similarity), there are difference in the information context (domain specific versus predominately universal) and the related achievable specificity and semantic relationships. Soergel (1999) argues that this perceived reinvention is "symptomatic of the lack of communication between scientific communities", resulting in ignorance of the

wealth of intellectual effort created by the library, information science and knowledge organisation communities. However, the 21st century and its drive for an inclusive Semantic Web has recently generated such communication between communities. W3C are recognising the value of existing controlled vocabularies (e.g. classification and thesauri), claiming that "OWL's precision is not always necessary of even appropriate" and "very strong outreach" to the library community, amongst others, is on the agenda (Herman, 2005). Based upon RDF, a new vocabulary has been developed to incorporate traditional knowledge organisation schemes and their associated collections into the Semantic Web (World Wide Web Consortium, 2005). Whilst OWL may be inappropriately specific, the Simple Knowledge Organisation System Core (SKOS Core) enables the interoperability of library classification schemes, representing their structure and content in a "machine-understandable form" (Miles et al, 2005). Developed with a recognition that "the semantic web is not just about interchange of data, but also about the *organisation* (my italics) of data in a distributed, decentralised way" (Miles et al, 2005), as will be demonstrated, SKOS Core is facilitating more scope for browsable interfaces. SKOS Core is currently an embryonic system, requiring further development and full standardisation (Miles et al, 2005) but it is an extremely welcome initiative to help bridge the gap between these previously disparate (yet common) communities. Furthermore, Powell (2005) believes that this enterprise may change the nature of our current traditional schemes, citing an example which is extremely pertinent to this current PhD research:

> If the whole of Dewey is available on the Web (with each concept having a unique URI) then I can easily pick and choose the bits I want and, more importantly, I can build extra bits around it.
>
> Powell (2005)

Such a vision may give Dewey the flexibility it needs to compete in the online world of the Semantic Web, with the deconstruction efforts described in this doctoral thesis being harnessed to fully realise this by facilitating a faceted view of the scheme. Dewey Decimal Classification initiatives are already underway, with researchers beginning to take advantage of this new simpler SKOS Core vocabulary, although not to use Dewey as the search tool, but to map a domain specific ontology onto the generic DDC (Li et al, 2005). These Canadian researchers have implemented this ontology mapping system to

enable end users to search for "learning objects" with their own domain-specific vocabulary (an ontology created in SKOS, based upon Canada's familiar provincial curriculum structure), which is then mapped (using SKOS Mapping) onto the eLera (2006) digital repository, classified by a modified version of Dewey. The SKOS Mapping between the different schemes takes the form of broadMatch, narrowMatch. exactMatch, majorMatch and minorMatch (Li et al, 2005; World Wide Web Consortium, 2004c). Preliminary results have been promising; in comparison to the previous restriction of just a keyword search, users can now conduct subject searches, retrieving a larger and more relevant result set. The subject search interaction is driven by users viewing terms in the domain ontology, listed by subject discipline and year group. The user then clicks a check box, alongside each term to include it in the search to be mapped. Thus, this is not a browsable system as such, allowing manipulation of a conceptual structure to broaden or narrow a search, but it does allow a user to view a series of subject terms, as opposed to specifying a subject search statement. Perhaps the major benefit of the system, though, is, as the authors conclude "the potential to exchange learning objects among repositories that use different classification systems" (Li et al, 2005). Thus, just as these researchers have mapped a domain-specific ontology as search tool to resources classified by the generic DDC, the reverse is, of course, possible. The future may see an online Dewey as search tool, mapped onto collections classified by domain-specific ontololgies. The task then is to provide more browsable system interfaces, which take full advantage of the conceptual structure of ontologies and classification schemes.

The focus, at present, however, seems predominately to be on the development of this 'internal' standardised metadata, hidden from the user (see also Vdovjak et al, 2003) , not how users might instruct and interact with agents and how search results might be presented and manipulated at the search interface. For example, Cimiano et al (2006) underpin British Telecom's Digital Library with an ontology, enabling end users to search using natural language questions, as opposed to conventional keywords; the end user, however, is not privy to the ontology's conceptual structure, which could drive an additional browsable interface. With the advent and increase of cross-database searching

and interoperability of ontologies and classification schemes, perhaps the next logical step will also include the exploitation of this encoded intellectual effort for conceptual display at the Web interface. As Fluit et al also claim:

> Many Semantic Web initiatives emphasize the capability of *machines* to exchange the meaning of information. Although their efforts will lead to the increased quality of the application's results, their user interfaces often take little or no advantage of the increased semantics. For example, an ontology-based search engine could use its ontology to enrich the presentation of the resulting list to the end user, for example, by replacing the endless list of hits with a navigation structure based on the semantics of the hits.
>
> Fluit et al (2006. p.45)

Although in addition to the presentation of results, as Fluit et al later demonstrate with the their discussion of the thesaurus-based Drug Onotology Project for Elsevier (DOPE) Browser (Aduna, 2005), a visual ontology can also play a key role in query formulation by displaying the entire conceptual structure of a domain and is associated documents, allowing the user to manipulate and refine the search space. Ontology visualisation for information retrieval within the Semantic Web is an "emerging research field" (Geroimenko and Chen, 2006. p.vi), with approaches and applications under development. One area of application has been the organisation of resources within web portals. For example, Miles et al (2005) cite the Semantic Web Environmental Directory (SWED, n.d.), underpinned by both OWL and SKOS Core and with terms extracted to create a facet-based interface:

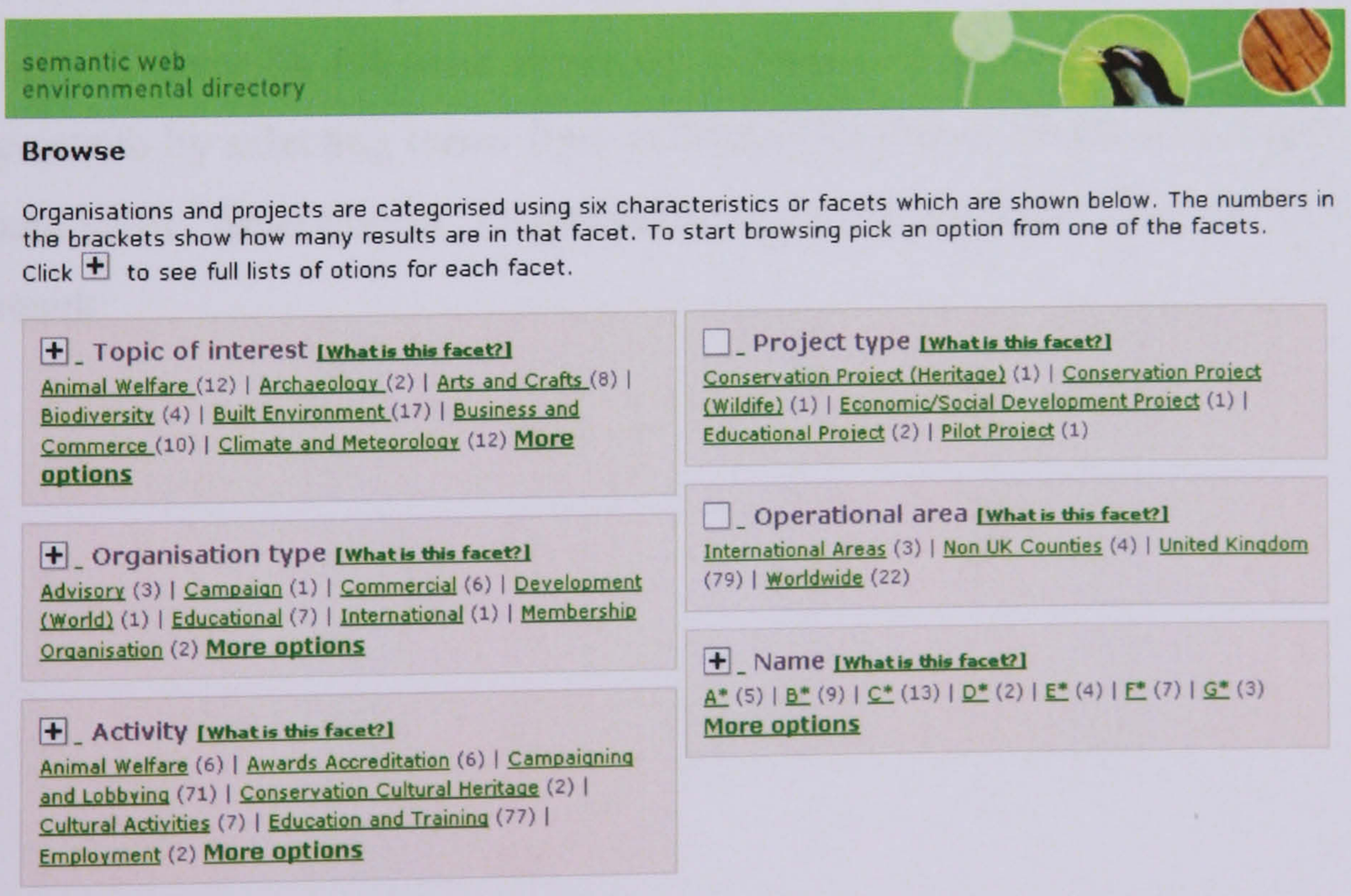


Figure 3.4: SWED Portal Opening Browse Interface

As shown in Figure 3.4, the interface contains six facets, with the number of results (web sites) alongside each term within the facet. The user can begin a search by selecting a term, for example 'Animal Welfare':

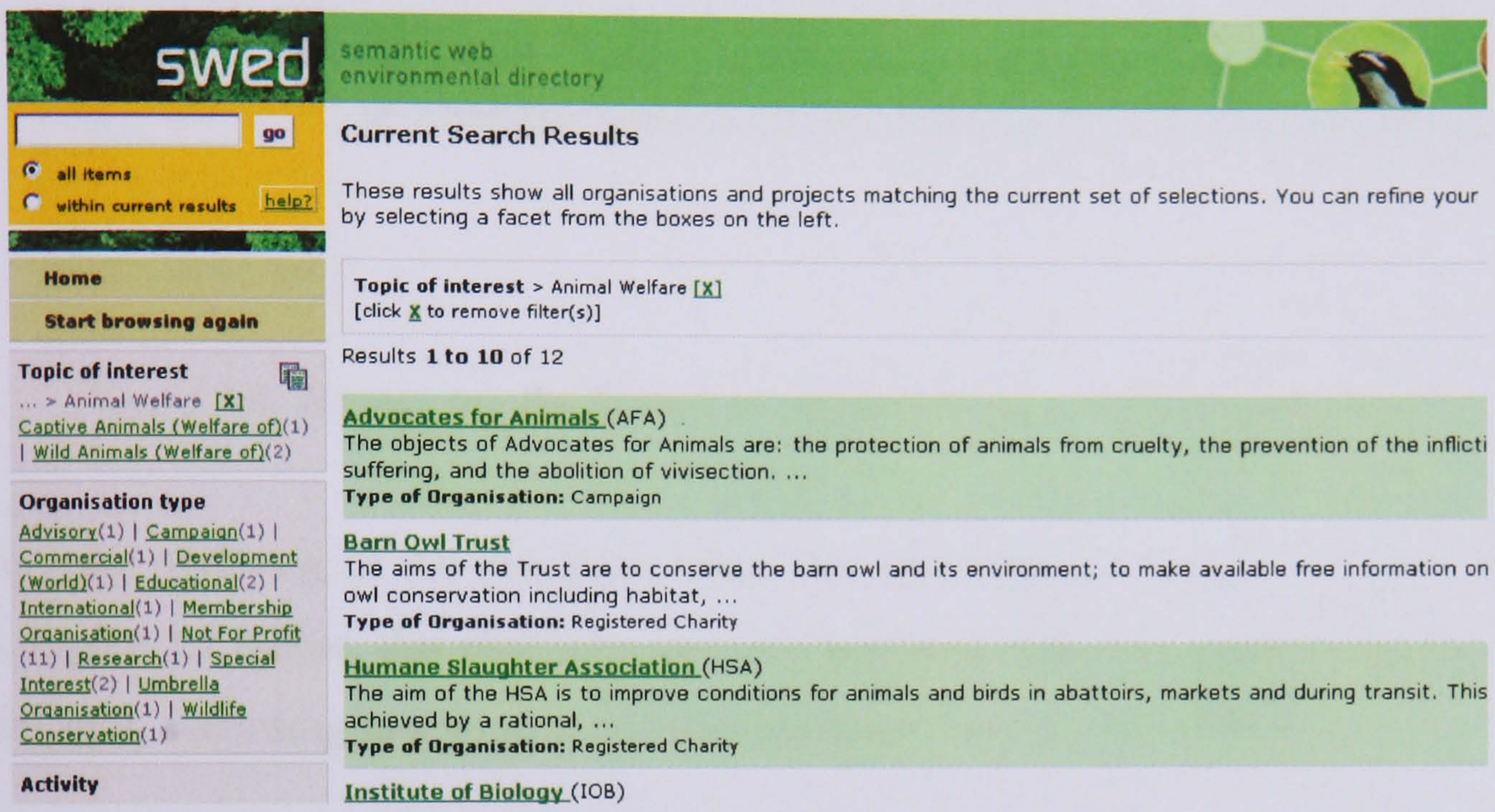


Figure 3.5: Results for 'Animal Welfare'.

As in our previously discussed View-based Searching (Pollitt et al, forthcoming), the Flamenco system (UC Berkley, n.d.) and the OPAC interface of this current doctoral research (see Chapter 8), the facet views are mutually constraining so the user can then refine the search by selecting terms from additional facets (seen here in a rather long list on left hand side). Selection of 'Organisation Type', 'Campaign' narrows the search to just one result:

swed

semantic web
environmental directory

go

all items
within current results help?

Home

Start browsing again

Topic of interest
... > Animal Welfare [X]

Organisation type
... > Campaign [X]

Activity
Animal Welfare(1) |
Campaigning and Lobbying(1) |
Education and Training(1) |
Information Related(1) |
Organisational Support(1) |
Wildlife and Biodiversity Related (1)

Project type

Current Search Results

These results show all organisations and projects matching the current set of selections. You can refine your by selecting a facet from the boxes on the left.

Topic of interest > Animal Welfare [X]
Organisation type > Campaign [X]
[click X to remove filter(s)]

Results **1 to 1** of 1

Advocates for Animals (AFA)
The objects of Advocates for Animals are: the protection of animals from cruelty, the prevention of the inflict suffering, and the abolition of vivisection. ...
Type of Organisation: Campaign

Figure 3.6: Narrowing the Search to 'Animal Welfare' AND 'Campaign'

A Boolean AND-ing of the two facet terms has restricted the search. As can be seen from Figure 3.7 below, this particular resource has been classified (almost indexed) very specifically, according to a number of different terms from multiple facets:

Advocates for Animals (AFA)

Type of Organisation: Campaign
URL:

The objects of Advocates for Animals are: the protection of animals from cruelty, the prevention of the infliction of suffering, and the abolition of vivisection. Advocates for Animals campaigns for parliamentary legislation to protect animals; finances and takes part in 'special investigations' into animal abuse; and supplies speakers and campaign literature wherever a request is made.

Advocates for Animals also seeks to make contact with those using animals, whether they are in the laboratory, on the farm or elsewhere, in an attempt to find common ground which might result in better welfare for the animals involved.

Founded in: 1912

Contacts

Name: name withheld (Director)
tel: 0131-225 6039
fax: 0131-220 6377

10 Queensferry Street
Edinburgh
EH2 4PG
Scotland

Classifications

Topic of interest	Animal Welfare | Education | Enquiries | Species | Teaching Resources
Activity	Animal Welfare SS | Awareness Raising SS | Campaigning and Lobbying SS | Information Provision SS
Operational area	United Kingdom SS

Figure 3.7: Record of 'Advocates for Animals' Web Page

This portal employs ontologies (not universal library classification schemes) and is concerned with a specific subject domain. However, it is presenting a faceted view of a more complex underlying ontology. This relates to the current PhD research in which library items are classified by a pre-coordinated Dewey notation with only explicit representation in one particular facet (subject discipline, the first component in the DDC code). The task is therefore to deconstruct the notations to represent a selection of fundamental embedded facets and the associated library items. This will allow a faceted view at the interface of the Dewey Decimal Classification Scheme to facilitate searching and browsing. This faceted approach to information retrieval is just beginning to receive attention within the Semantic Web community. Another team of Semantic Web researchers (Hyvönena et al, 2004a; Hyvönena et al, 2004b) are marrying View-based Searching with ontology metadata at an image search interface:

> The underlying domain ontologies are projected into facets that facilitate multi-facet search
> Hyvönena et al (2004a)

This multi-facet search is then supplemented with a "recommendation system", which can suggest related images as hypertext links, based on the user's initial search and underpinned by a set of knowledge-based recommendation rules. The Ontogator system has recently been implemented in the form of a portal to the Museum of Finland (see Semantic Computing Research Group, 2004). The promotion and exploitation of such underlying classification-based ontologies to an explicit role at the user-interface, most notably within a facet-based search, is beginning to emerge and appears extremely promising for future Semantic Web research. In a recent paper, perhaps reminiscent of the seminal declaration of the Classification Research Group in 1955 (see Chapter 2, p.27), Mäkelä et al (2005) proclaim "the multi-facet search paradigm as a basis in information retrieval on the Semantic Web." So, as previously discussed, although enumerative, hierarchical classifications continue to dominate in the printed world, perhaps the expressive power and flexibility of faceted classification will be recognised and realised in another: a world of online information retrieval and the Semantic Web.

Similar recognition of the value of knowledge structures for information retrieval at the OPAC interface, however, is still sorely lacking.

In summary, therefore, in contrast to commercial library OPACs, the Internet and online database communities have for some time been exploring the use of 'home-grown' and standard classification schemes for subject browsing, as an alternative or supplement to keyword searching. The prominent approach is hierarchical classifications, although typically applied in a rather linear, shelf-oriented manner. However, certain researchers and knowledge management specialists are beginning to discover the potential and flexibility of faceted classification for providing a more multidimensional access to information resources. Allied to this is the recent development of the Semantic Web, which, although not predominantly utilising classificatory structures at the system interface as yet, is recognising the value of its inherent vocabulary control to create structured metadata for enhanced Internet searching. Although initiatives to exploit underlying ontologies at the user interface are currently relatively few, these few may begin to set a precedent and preference for faceted retrieval.

3.8 Conclusion

This chapter has charted the development of the library catalogue, from printed to online access, and the changing role of classification within this fundamental information retrieval tool. In early catalogues, even in name, classification was given a prominent role; across the continents, the classified catalogue rivalled the dictionary catalogue. Yet, later, as standards were sought, the dictionary catalogue became the predominant model and the role of classification as an information retrieval tool diminished, relegated to a 'mark and park' tool for shelf location. Furthermore, this diminished role was then underwritten by the emergence of computer technology - in which systems were designed for the highly-trained expert not the end-user - followed by a library OPAC, modelled on such batch-mode systems and the foregoing 'known-item' dictionary card catalogue, with no provision for subject searching, unlike the classified catalogue model.

An OPAC borne from a dictionary catalogue and exact-match query specification model thus prevailed, despite growing research indicating that users were experiencing subject search problems. Researchers called for library systems that could help users explore and discover information, not simply find that already known. Conceptual displays, such as classificatory structures, were cited as a possible method for achieving this kind of access. However, despite several experimental systems, exploring and providing evidence of the benefits of classificatory structures at the interface, little change emerged from the commercial OPAC domain.

It was not until the Internet community began to rediscover and reorient a classificatory approach within search engines, coupled with elements of standard classifications being explored by the library community within various subject gateways that classification for information retrieval re-emerged. Linear hierarchical approaches have tended to dominate, although multidimensional, facet-based, approaches are beginning to appear, together with the emergence of classification-based metadata in the form of the Semantic Web and associated faceted interfaces. Library OPACs, however, still adhere to their specification-based form-fill-in interfaces, with little exploitation of the classification

schemes therein; whilst libraries still maintain a predominantly hierarchical/enumerative classificatory approach in the physical domain, viz. the Dewey Decimal Classification scheme, as described in the previous Chapter 2. Thus, it would be difficult to apply the structures of a faceted classification in an online domain, as described in this chapter, to a physical collection already classified according to a hierarchical/enumerative scheme; we have to take account of the current context and work within, yet aim to improve, this.

If we are to enhance subject searching and improve access to physical collections via online applications of their classificatory structures within library OPACs, whilst avoiding the restrictions of shelf-based applications of hierarchical classification schemes observed in Internet applications, then we should look at *exploiting the faceted potential within such hierarchical/enumerative schemes.* This will be explored in the forthcoming chapters and will not only move such online hierarchical applications beyond unnecessary replications of physical linear shelf arrangements, which is currently at odds with the multidimensional nature of today's information and library items and the diverse needs of users, but will exploit the immense intellectual effort lying dormant in our library OPACs.

CHAPTER 4

METHODOLOGICAL APROACH AND DECISIONS

> For the most part twentieth-century library classification has subscribed to the idea that somewhere, somehow, we can, or should try to, produce the one best classification system that will serve all purposes...The fact is, however, that knowledge organization systems appear to be inherently personal, and while many people in groups might agree in great part on how such systems should be structured, there is in the end no one best classification of knowledge system in an absolute sense. There are only *alternative arrangements* (my italics) that serve one or another purpose or person for better or for worse.
>
> Miksa (1998, p.81)

4.1 Introduction

Classification in the physical world is always a compromise. Attainment of the most 'helpful arrangement' for the majority of users is the culmination, and most visible result, of the classificatory process in a physical domain. However, fulfilling this is now more and more difficult, as the subject content of information items is becoming increasingly multi-faceted and the information needs of users more diverse. One size cannot, and ideally should not, fit all. Classification in an electronic domain creates a potential for multiple locations and alternative arrangements, a flexibility freed from the constraints of physical access. However, such exploitation and manipulation of the faceted nature of classificatory structures, readily available within library OPACs for searching and browsing, still remains untapped.

This chapter focuses on the Dewey Decimal Classification scheme, considering the problems intrinsic to the classified arrangement of physical library collections and how subject access could be enhanced through alternative arrangements in the online domain, based on the hidden embedded facets within library item DDC notations. The chapter forms a preface to the procedural method detailed in the following Chapter 5. Firstly, the problems created by the need for a citation order in the physical classificatory domain are demonstrated; secondly, it is explored how such problems could be addressed by deconstructing Dewey into its component facets to provide alternative arrangements; finally, the methodological approach to deconstruction employed in this current thesis is introduced and explored.

4.2 The Effects of Citation Order

It was discussed in Chapter 2 how a Dewey notation can contain multiple subject components, which can be ready-built (enumerated) within the schedules or 'built' (synthesised) by re-using schedule code components or notations from the now six auxiliary tables. When building any number, a citation order comes into play, reflecting the physical basis of a scheme primarily intended for shelf organisation. The term 'order' implies placing documents in a sequence, necessitating a single ordering principle based on a designated primary characteristic; however, documents can be about multiple topics, and moreover, these distinct topics themselves might also merit treatment from different viewpoints. For instance, the topic of 'glass' could be treated in relation to the decorative arts; however, this concept could also be viewed from and related to the perspectives of chemical engineering and building construction, for example. Although classification recognises and makes provision for representing the multidimensionality of the universe of knowledge, problems arise when this is translated into the physical domain of shelf arrangement, whereby an item is restricted to a single location in a linear sequence.

Thus, the consequence of a citation order is that only the first element (primary subject) of a multi-faceted library item is represented explicitly in the shelf order, leaving the remaining subject facets undiscovered. Such hidden facets may be fundamental to certain users yet do not take precedence in the citation order and, hence, the shelf arrangement; for instance, a user may want to identify library items about a particular country or time period, regardless of disciplinary subject matter. The richness and multidimensional nature of many library items is often obvious from its title or Dewey code. For example, a library item entitled *The* **Lancashire** *cotton industry : a study in economic development,* having the Dewey code 338.4767721094276 has the following hierarchical structure:

The Lancashire cotton industry : a study in economic development

Assigned DDC Code: 338.4767721094276

Schedules	**3**	Social Sciences
↓	**33**	**Economics**
	338	Production
	338.1 – 338.4	Specific kinds of industries
	338.**4**	Secondary Industries and Services
	338.47	Goods and Services
Built from Schedules	338.471 – 338.479	Subdivisions for Goods and Services
	338.476	Technology
↓	338.4767	Manufacturing
	338.47677	Textiles
	338.476772	Textiles of Seed hair fibres
	338.4767721	**Cotton**
Built from Table 1	338.47677210	Facet Indicator for Standard Subdivision
↓	338.476772109	Historical, geographic, persons treatment
	338.4767721094	Europe Western Europe
Built from Table 2	338.47677210942	England and Wales
↓	338.476772109427	Northwestern England and Isle of Man
	338.4767721094276	**Lancashire**

Figure 4.1: Hierarchical Structure of a Multi-faceted Library Item

Figure 4.1 not only demonstrates the inherent richness and subject specificity capability of Dewey notations, but also how the three facets (economics, cotton, Lancashire) represented in the library item title are encoded in the Dewey notation. As stated in Chapter 2, the DDC divides and organises the world's knowledge by subject discipline, whilst our universe of knowledge becomes more inter-disciplinary. Thus, in Figure 4.1, the subject discipline of economics takes precedence over the other subject facets, and consequently is the first notational component in the citation order and thus the only one made explicit to the user in the library shelf arrangement. Facets such as 'Place' (here, 'Lancashire') are relegated to the end position in the code. As a result, such embedded facets are scattered across the main class Dewey disciplines, as exemplified by the following table of library items:

DeweyCode	Title
304.20942769	Restructuring : place, class and gender
305.56094276	Working class Barrow and Lancaster 1890 to 1930
307.2094276	Migration and social adjustment : Kirkby and Maghull
331.110942766 5	The dynamics of working-class politics : the labour movement in Preston 1880-1940
331.11109 4276	Jobs in the countryside : studies in north Lancashire and Cumbria
333.7094276	Public perceptions and sustainability in Lancashire : indicators, institutions, participation
338.4767721094276	Lancashire textiles : a case study of industrial change
338.476772109 4276	The Lancashire cotton industry : a study in economic development
362.210942765	The desegregation of the mentally ill
363.2094276	Law and order in early Victorian Lancashire
365.7094276	Prison reform in Lancashire, 1700-1850 : a study in local administration
374.01309427642	Report by HM Inspectors on Burnley College
574.50942764 2	The natural history of the Burnley area
711.3094276	Tuning in to the public : survey before participation
711.4094276	Central Lancashire New Town proposal : Impact on north east Lancashire[Consultants' appraisal]
711.4094276	North East Lancashire structure plan : public participation
711.409427612	Skelmersdale New Town planning proposals : report on basic plan
728.30942764	Rural houses of the Lancashire Pennines 1560 to 1760
728.64094276	Handloom weavers' cottages in central Lancashire
792.0942769	The Grand Theatre, Lancaster : two centuries of entertainment

Table 4.1: Scatter of 'Lancashire' (-4276) across the DDC Schedules

Table 4.1 shows how 'Lancashire', and its more specific localities (indicated by further digits, coloured blue), is distributed across three main Dewey classes: *300 Social Sciences, 500 Natural sciences and mathematics* and *700 The arts Fine and decorative arts.* Similarly, Figure 4.2 illustrates the conceptual distance and Figure 4.3 the physical distance of this scatter.

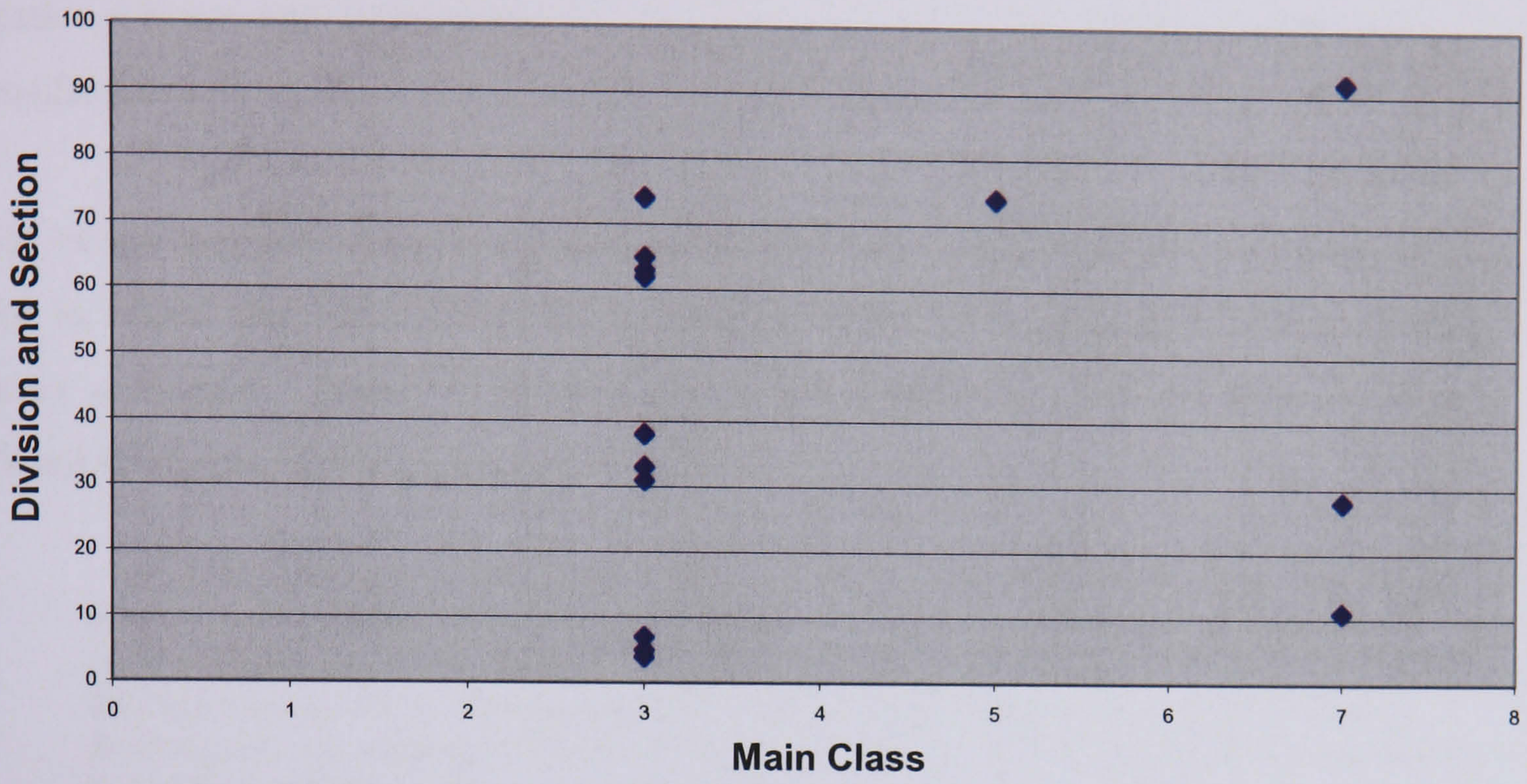


Figure 4.2: Graph illustrating the Conceptual Scatter of 'Lancashire' across the Dewey Schedules

Dewey Section Code	Dewey Caption	Library Floor Location
304	Factors affecting social behavior	Floor 6
305	Social groups	
307	Communities	
331	Labor economics	Floor 5
333	Economics of land and energy	
338	Production	
362	Social welfare problems and services	Floor 4
363	Other social problems and services	
365	Penal and related institutions	
374	Adult education	Floor 3
574	Biology (20th ed.) now unassigned	
711	Area planning (Civic art)	
728	Residential and related buildings	
792	Stage presentations	Floor 2

Figure 4.3: Physical Scatter of 'Lancashire' across the University of Huddersfield Library Building

The physical arrangement and display of library items, then, can only represent one viewpoint, the choice of which is determined by precedence and that which satisfies the majority of user needs: the most 'helpful order'. This 'helpful order' is typically on the

basis of subject discipline, with topics such as 'place' and 'time' (unless they are, of course, the primary topic of study, namely geography or history) constituting secondary topics. Choice and compromise is therefore implicit in any physical application of classification.

This current research is therefore seeking to collocate library items by embedded facets such as 'Place' and, in so doing, derive alternative Dewey arrangements of a particular library collection. Miksa (1998) underlines the importance of such alternative user-tailored arrangements in his future vision for the DDC:

> ...to achieve this kind of malleability, we will have to view the structure of the DDC differently than we have in the past. We will have to see the entire system as a vast array of moveable or interchangeable facets of categories, a system which is perhaps best called an object relational database management system of categories...If the DDC continues on its present course of development - i.e. adopting the library classification theory and techniques that evolved in the first half of the twentieth century - is its future assured?
>
> Miksa (1998, pp. 89-90)

The advantages afforded by the electronic domain for Dewey manipulation were recognised by Wajenberg in 1983. He recommended that library item Dewey notations should have their component parts encoded in OPAC MARC records, revealing their hierarchical structure (such as that of Figure 4.1) and means of construction. It was envisaged that the adoption of such cataloguing practice would lead to an increase in the number of subject access points, enabling a user to search for a particular element embedded within a Dewey code and its collocation across other composite notations. Wajenberg exemplifies this as follows:

> If a student of comparative literature wished a comprehensive listing of works about symbolism in poetry, the system could search not only the class number 809.1915 (the number for general and comparative studies of symbolism in poetry), but also every occurrence of 1009 from Table 3, followed by 15 from DDC Table 3-A. That would retrieve works on symbolism in French poetry, American poetry, etc., wherever those tables had been applied.
>
> Wajenberg (1983, p.251)

Similarly, the first recommendation arising from the OCLC Dewey Decimal Classification Project (Markey & Deymeyer, 1984) was for "the identification and implementation of a coding scheme for the individual elements of a synthesized class number in libraries' bibliographic records" (p.323). As highlighted in Chapter 3, the

Dewey Online Catalog (DOC) was the first to implement the Dewey Schedules and Relative Index in a library OPAC for searching and browsing. However, the DOC system was, at Markey and Deymeyer's own admission, limited to a best match of an *enumerated* (i.e. 'un-built') Schedule notation, thus excluding a substantial number of Dewey notational combinations. They recommended that Wajenberg's coding scheme be adopted, evaluated and revised so as to include all methods of Dewey number building. This would entail indexing of synthesised notation components and therefore incorporate access to the Dewey Tables (a series of notations used when building/synthesising a class number). However, perhaps due to cost, libraries never embraced this recommendation.

As discussed in the previous Chapter 3, even in today's commercial library OPACs, the Dewey code is still used primarily as a shelf-locating device, merely the culmination not the means of an OPAC search. If utilised at all for searching, it can only be accessed in its entirety or, at best, be truncated from the first notational component, i.e. there is no means of searching on the embedded subject components. Moreover, motivation to utilise Dewey codes for searching assumes recognition by the user that the notation encodes and collocates subject information; however without the accompanying Schedule captions and hierarchical structure, this remains inexplicit. In an exploratory observational study of OPAC use in the University of Huddersfield Library only 1 in 68 users chose to conduct a search via a Dewey code (Tinker, unpublished report, see Appendix 1). Consequently, despite the potential embodied in the online domain of commercial library systems, usage of the DDC is still confined to its 'mark and park' role and, if utilised for searching at all, requires a user to enter a known notation for subsequent entry into a index listing of library items titles, replicating the linear order of the shelf arrangement.

4.3 Revealing Embedded Components via Decomposition

Therefore, since the Dewey code components are already pre-coordinated in such library systems, the only means of providing enhanced subject access via such embedded subject facets is by decomposing (deconstructing) Dewey notations into their component parts; in effect providing an additional post-coordinate view of the DDC.

n. Decomposition - The action or process of decomposing, separation or resolution (of anything) into its constituent elements

Oxford English Dictionary

In order to decompose Dewey, one must first examine the rules of its construction. To reiterate briefly from Chapter 2, a Dewey classification code can be assigned to an item using a ready-made notation enumerated in the Schedules. This notation may already signify more than one facet (e.g. 032 General Encyclopedic Works in English, 061.1 General Organisations in Canada). Additional facets can be appended by re-using Schedule code components or notations from the six auxiliary tables. A Schedule base number can be extended by another full class number or part of one. This notational synthesis is applied either under instruction (Individual and Collective 'Add notes') in the Schedules or without instruction using the first auxiliary table, the Standard Subdivisions. Liu (1993) has previously undertaken such analysis and decomposition with considerable success (100%, excluding incorrectly constructed Dewey notations). Liu analysed and encoded 17 rules used for synthesising Dewey notations, building a system (the Dewey Number Decomposer) that decomposed a test database of 6000 codes within the Dewey Main Class '700 The Arts'. It was claimed that this class, excluding Language (400) and Literature (800), is representative of the other Dewey classes and thus is readily extendable; however, others never pursued this. Moreover, decomposition efforts have not been restricted to the DDC. Gerhard Riesthuis (1997) has undertaken research using a similar rule-based approach to formulate algorithms for the decomposition of complex Universal Decimal Classification (UDC) notations.

The approach to decomposition described in this thesis is more end user-oriented than that of Lui (1993, 1996) and Riesthuis (1997), entailing the derivation of several fundamental facets, applicable to any subject discipline and thought to provide useful subject search views at the OPAC interface.

4.4 Identifying and delineating facets

Ranganathan's concept of fundamental facets was a source of inspiration when deciding upon which particular facets would underpin alternative views of the Dewey Decimal Classification Scheme, as described in this thesis. Thus, unlike Liu's (1993, 1996) work, instead of decomposing a single subject-specific class, this current research adopted a holistic approach, focusing on identifying generic facets which may be helpful to a user from *any* subject discipline. The task, therefore, was to identify common classificatory components, reused or applicable throughout the scheme.

Facet identification is difficult in Dewey, relative to other classification scheme notations. For example, as outlined in Chapter 2, Ranganathan separated each facet using a colon, hence the name of the scheme; similarly, the UDC uses symbols such as +, :, =, ( ) etc. to denote facet boundaries and the nature of the following facet (e.g. language, form, time etc.). Facet indicators appear but are not integral to the DDC, the main instance being a zero to introduce Standard Subdivisions (the main source of fundamental facets), although their use in Life Sciences and public administration Schedules are recent additions. However, the use of zero does not consistently indicate a subsequent Standard Subdivision notation, as Wajenberg (1983, p.247) recognised:

> Every Standard Subdivision number is preceded by at least one zero. Unfortunately, however, that is not the only use of zero in the DDC, so it cannot be used as a signal to a computer that what follows is a number from the DDC Table 1.

Allied to this is the lack of a uniform notation; the same notation may be used to denote several different concepts across the classification scheme; or, viewed conversely, the same concept can be represented by different or distorted notations, the latter caused by partial re-use of notational components when number building. Miksa endorses alternative arrangements based on 'collocations of categories' in his future vision for the DDC but, as shown by the inconsistencies below, it would be difficult to collocate on notation:

It would require something akin to a change in template for moving an entire series of collocations from one place to another on the initiative of the DDC user. Thus, if a classifier wished to collocate a wide variety of information sources (or electronic links to them) related to jewelry, the template should allow the user to move the topical areas and subclasses of, say gem metallurgy (553.8), gem mining (622.38), synthetic gems (666.88), gem carving (736.2) jewelry and gems in religion (the categories of individual religions from the 270s to 290s), and so on into a structure at, say, 739.27 (Jewelry) with links from their standard positions to the new locations.

Miksa (1998, p.89)

The Editor-in-Chief of the DDC, Joan Mitchell (1997, p.86), has pointed to the value of a standard notation for information retrieval, namely for resolving disciplinary scatter and enabling decomposition and collocation. She cites the use of standard notation for bears (my use of colour):

> The use of a standard notation for bears following the facet indicator provides hierarchical links to broader concepts (carnivores 599.7, mammals 599) and narrower concepts (grizzly/brown bears 599.784, American black bear 599.785, polar bear 599.786, and giant panda 599.789). By using uniform notation to represent bears across the Classification, we are able to retrieve bears in the sense of carnivorous mammals in may disciplines (e.g. conservation of bears [639.97 + 978) and not retrieve bears in the sense of teddy bears (toys):
>
> | | |
> |---|---|
> | Bears | 599.78 |
> | animal husbandry | 636.978 |
> | big game hunting | 799.2778 |
> | technology | 639.97978 |
> | resource economics | 333.95978 |
>
> and not
>
> | | |
> |---|---|
> | Teddy bears | |
> | commercial manufacture | 688.7243 |
> | handicrafts | 745.59243 |

The notational re-use in the above example initially looks extremely promising for collocating concepts across subject disciplines; however if we search across the Dewey Schedules for the --978 notation, it can be seen that its use is not exclusive to 'bears':

Dewey Class Code	Caption
299.78	Religions of specific groups and peoples
343.0975-343.0978	Air transportation
343.0978	*Air transportation services
469.78	Geographic variations in Madeira
491.59701-491.5978	Subdivisions of Kurdish (Kurmanji)
549.78	Carbonates
549.782	Calcite group
549.785	Aragonite group
571.978	!Tumors
599.78	***Ursidae (Bears)**
599.784	***Ursus (Grizzly bear, Brown bear)**
599.785	***Eurarctos (American black bear)**
599.786	***Thalarctos (Polar bear)**
599.789	***Ailuropoda (Giant panda)**
616.978	*Autoimmune diseases
618.920977-618.920978	Ophthalmology, otology, audiology
618.976-618.978	Specific diseases
621.978	Riveting equipment
639.972-639.978	Specific kinds of animals other than mammals
745.61978	Roman calligraphy
891.59701-891.5978	Subdivisions of Kurdish (Kurmanji) literature
939.78	*Ethiopia to 500
978.401-978.403	Historical periods

Table 4.2: Re-use of -978 notation to represent different concepts

Therefore, although a uniform notation may exist for the particular concept of 'bears' if examined in isolation, i.e. no other notation is used, it is not used solely for this concept and thus any attempts at decomposition and collocation cannot reliably be undertaken based on notation alone without retrieving 'false drops'. There is, however, a move towards the 'regularization' of Standard Subdivision concepts, removing irregular enumeration of such concepts in the Schedules and replacing it with the consistency of standard notations from Table 1.

The general lack of a uniform notation placed restrictions on the scope of the decomposition efforts described in this thesis, directing the research away from the identification of particular cross-disciplinary subject facets, as implied by the examples of Miksa and Mitchell, and towards the identification of fundamental facets, derived from

the more consistent notations of Table 1 (Standard Subdivisions). The Standard Subdivisions can be seen as having the following properties, conducive to decomposition:

- *Recurring non-primary characteristics of a subject*
- *Non-topical characteristics that pertain to the document itself rather than to its primary subject.*
- *Applicable to any class number for any topic that approximates the whole of the number.*
- *With a few exceptions, attached to any enumerated or built number without formal add instructions.*
- *Consist of at least 2 digits of which initial is zero - constant feature and was first use of facet indicator in library classification. Marks transition from primary subject to secondary subject.*

Figure 4.4 Properties of Standard Subdivisions[9]

The non-topical nature, general applicability and relative notational consistency of the Standard Subdivisions therefore made them the ideal source for deriving and populating facets, the process of which is described in the following chapter.

The Standard Subdivisions were therefore the basis for facet identification due to the above properties, enabling their relative ease of extraction. It was also the intention to collate library items which were subject to high level of disciplinary scatter; this is typified by the generic nature of Standard Subdivision components. Another consideration was the desire to select facets that would be relevant to a diverse range of users from any subject discipline, provided for by the generic Standard Subdivisions, which can be added to any Schedule notation unless instructed otherwise. A strong literary warrant within the University of Huddersfield (yet also within the majority of

[9] Summarised from Chan et al (1996, pp. 97-98).

academic library collections) was another factor; this would enable effective facet population by targeting and collating large subsets of library items, which may constitute useful subject access points for a wide range of users across disciplines.

Based upon these considerations four candidate facets were selected as case studies:

- *Form/'Reference Material Type'*

Key reference materials (dictionaries, encyclopaedias, biographies, bibliographies) which, unless generic (located within the Dewey Generalities Main Class), are buried and scattered across and within subject disciplines

- *Persons* – Women, Men and Children

This was selected with a view to collating interdisciplinary subjects such as 'Women's Studies', which consequently suffer from a very high level of disciplinary scatter. It is also anticipated that the other categories (Men and, particularly, Children) will also provide useful subject access points for users.

- *Time*

This concept has a strong literary warrant across many subject disciplines and hence is potentially relevant to a substantial proportion of users. 'Time' is currently embedded as a secondary facet within Dewey notations across subject disciplines, yet also as a primary facet within subjects such as history. Both generic time categories by century (added via the Standard Subdivisions) and enumerated time periods by geographic location coexist; the task is therefore to reconcile these different approaches and library item disciplinary scatter within a single, additional 'Time' Facet.

- *Place*

Like 'Time', this facet has a strong literary warrant and high levels of disciplinary scatter. It is similarly applied both generically, as a secondary concept via the Standard Subdivisions and Table 2, and subject specifically via enumeration as a

primary facet within the Dewey Schedules. Again, the task is to collate these scattered library items within a single Place Facet.

Further rationale for the specific choice of each facet is provided in chapters 6 and 7.

4.5 Conclusion

This chapter has examined the problems and compromises of library classification, when applied in the physical domain of shelf organisation, and how such problems might be approached and addressed. The Dewey Decimal Classification Scheme organises knowledge according to subject discipline and a linear pre-requisite citation order, which collects by the first classificatory component whilst scattering by the remainder. This may cause problems for users wishing to search or view a library collection according to alternative arrangements, by non-primary facets such as time and place. Although such alternative classificatory arrangements are precluded in the physical domain, classification schemes can be manipulated in an electronic environment. However, whilst commercial library OPACs have failed to exploit the structures of classification, let alone its malleability, a few researchers have investigated decomposing a classification to enhance subject access using rule-based methods. However, due to inconsistencies within the DDC, the current research will adopt an alternative method. This method entails the identification of fundamental facets, based upon embedded Standard Subdivision components and then populated with relevant library items, to create alternative arrangements of a physical library collection via the OPAC interface. Four facets (Form/Reference Material, Person, Time and Place) were identified, briefly outlined and their choice justified. The methodological procedure is described in the following Chapter 5 and explored further by the discussion of case studies in Chapters 6 and 7.

CHAPTER 5

PROCEDURAL METHOD: DERIVING AND POPULATING DEWEY FACET VIEWS

5.1 Introduction

The aim of the empirical work described in this thesis is to derive facet views of the Dewey Decimal Classification Scheme for application in a View-based Searching library OPAC. Following the methodological decisions and approach of the previous chapter, this chapter now describes a procedural method devised and utilised to accomplish this task. The current chapter firstly details the components, followed by sections outlining the procedure of facet derivation and population. It should be noted that this chapter presents only an overview of the method; a more detailed and illustrated description and analysis is given within the context of two case studies, examined in the subsequent two chapters.

5.2 Components

5.2.1 Dewey Decimal Classification Scheme

The Schedule codes and captions of the Dewey Decimal Classification Scheme (21st edition) were imported into a Microsoft Access database. The hierarchical structure of Dewey, together with the additional derived facets, will ultimately form the meta-data and information retrieval tool utilised for searching and browsing a library collection in the created OPAC system.

5.2.2 Test Database

The Dewey codes, titles and publication year of 255, 212 library items, comprising the entire collection of the University of Huddersfield library, were downloaded into the same Microsoft Access database.

5.2.3 SQL Select Queries

Once identified, the facet views were populated with library items by way of 'pattern matching' select queries (in certain cases, assisted by positional criteria/facet citation order), utilising Dewey code and, on occasion, library item title components as query criteria.

5.2.4 Search Interface

Following derivation and population of the facets within the Microsoft Access database, the results were incorporated into a View-based Searching application (described and illustrated in Chapter 3).

5.3 Procedure

The following sections describe the procedural method; a graphic representation of which is given in Figure 5.1, highlighting a facet derivation and population process that combines manual intervention with the more automatic SQL. Figure 5.1, the initial facet identification is manual, (coloured grey) yet queries (coloured blue) are then manually created that enable the automatic identification of numerous relevant library items, which will populate the facet. Having initially identified these library items, the results are manually evaluated (checked) for possible false drops and, where appropriate, the query criteria is further refined and then the query re-run. Similar manual intervention then ensures that the library item results correctly map onto the facet table. Where appropriate, a translation table may be manually created to enable an exact match or a string expression is incorporated into the query to extract the matching component. Finally, the results are combined by a union query to populate the facet. The following sections describe this procedure in further detail.

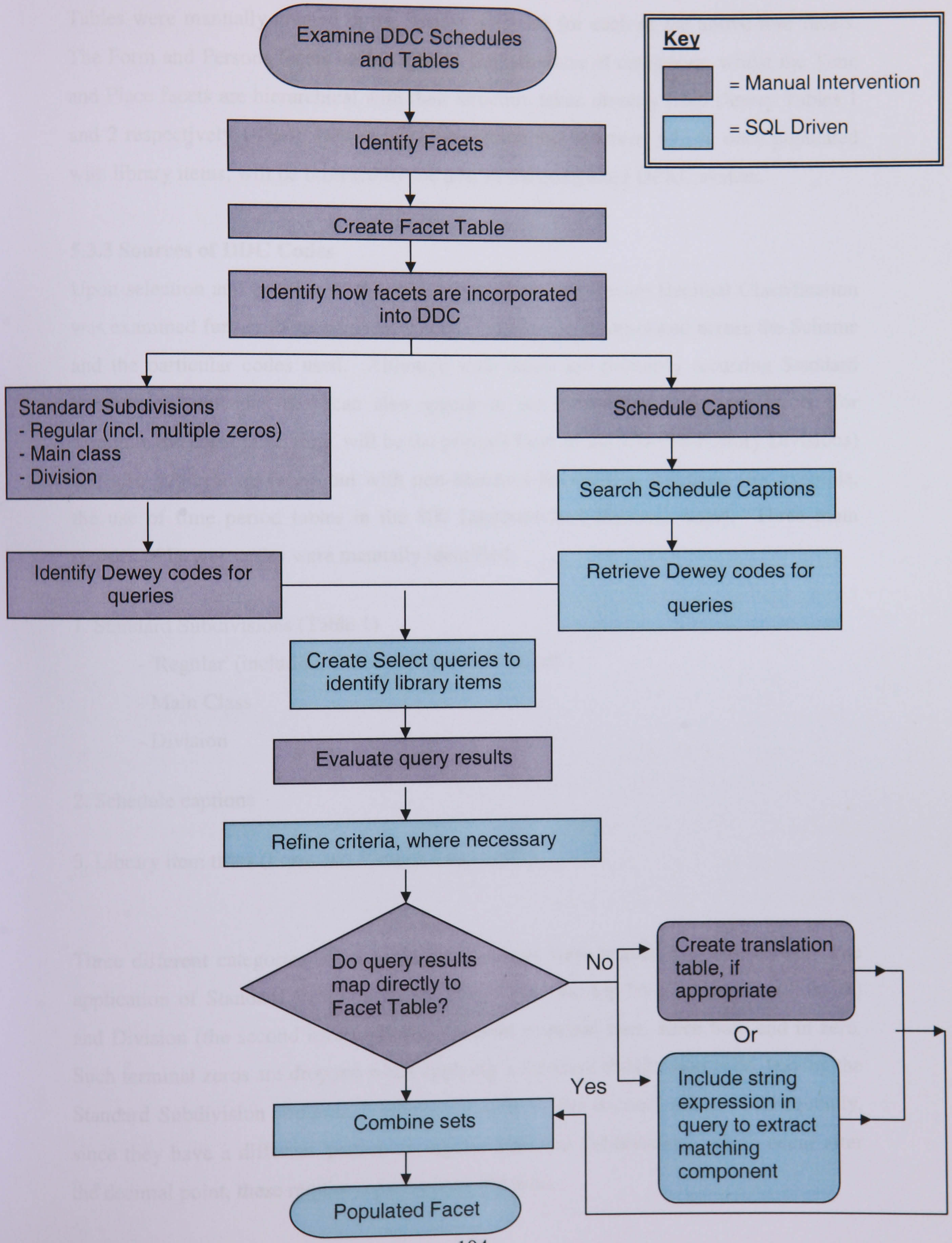
5.3.1 Facet Identification

The Dewey Decimal Classification (namely its Schedules, Tables and rules for synthesis). coupled with literature about the scheme, were studied to ascertain the extent of the task and possible approaches to facet derivation. As stated in Chapter 4, Standard Subdivision (Table 1) notational components, due their relative consistency, facet indicators and cross-disciplinary applicability, were chosen as the basis for identifying facets. As detailed in the previous Chapter 4, four facets were manually selected as case studies:

- Form/'Reference Material Type'
- Persons
- Time
- Place

The derivation and population of each of these is discussed in detail in the following chapters.

Figure 5.1: Outline Procedural Method for Dewey Facet Identification and Population



5.3.2 Facet Creation

Tables were manually created in the Access database for each of the above four facets. The Form and Persons facets have a simple 'flat' structure of categories, whilst the Time and Place facets are hierarchical with their structure taken directly from Dewey Tables 1 and 2 respectively. These Dewey tables constitute the structure, which, once populated with library items, will be browsed by the user in the completed OPAC system.

5.3.3 Sources of DDC Codes

Upon selection and creation of the above four facets, the Dewey Decimal Classification was examined further to ascertain how such concepts are incorporated across the Scheme and the particular codes used. Although such facets are primarily recurring Standard Subdivision concepts, they can also appear in the Schedules as primary facets (for example, the concept of 'time' will be the primary facet in the 930-990 History Divisions) and also subordinate facets but with non-Standard Subdivision notations (for example, the use of time period tables in the 800 Literature and Rhetoric class). Three main sources of Dewey codes were manually identified:

1. Standard Subdivisions (Table 1)
    - 'Regular' (including codes with multiple zeros)
    - Main Class
    - Division

2. Schedule captions

3. Library item titles (Form and Persons Facets only)

Three different categories of Standard Subdivisions were manually distinguished. The application of Standard Subdivisions to Main Class (the top level of the classification) and Division (the second level) notations present a special case, since both end in zero. Such terminal zeros are dropped when applying a Standard Subdivision code, leaving the Standard Subdivision component effectively split by the decimal point. Consequently, since they have a different 'pattern' to regular Standard Subdivisions, which occur after the decimal point, these require separate identification.

5.3.4 Identifying Library Items

Select queries were created to retrieve library items for each of the four facets in turn, based on the above three sources of Dewey codes. Due to the lack of a completely uniform notation and the re-use of DDC codes for different concepts, the construction of queries was on occasion a recursive process, requiring refinement of query criteria to exclude false drops.

Identifying library items with Standard Subdivision components entailed utilisation of the zero facet indicator, for example '--09' to identify library items with a place facet. The Schedule captions were also searched for relevant concepts, retrieving a list of Dewey codes, which were subsequently incorporated into queries to match library items. Finally, library item titles were utilised for the Persons and Form facets. The restriction of title searches to just these two facets was due to the fact that both contain only three and four categories respectively, each of which were dealt with by separate queries. In contrast, the Time and Place facets contain numerous categories and, similarly, would require multiple title searches. Consequently, this entailed one-to-one matching of Dewey code to library item as opposed to the many-to-many matching of the Time and Place facets.

5.3.5 Matching Results to Facet Table

Having identified relevant library items, there is a need to convert or assign a code equivalent to that used in the Dewey facet table. The Form and Persons facets are 'simple' in that library items for each category within the facet are identified with separate queries, entailing a one-to-one matching of Dewey code to library item. Dewey codes are not used in the Dewey tables of these 'simple' facets, instead three letter ID codes identify each category, e.g. bibliographies in the form facet are coded BIB, biographies BIO etc. These facet ID codes were incorporated into the previously created queries to identify and now appropriately code library items so as to match their respective facet tables.

The same procedure is more complex for the Time and Place facets, since Dewey notations structure the facet table. However, the library item Dewey code in its current

form will not match the facet table, since the requisite component is embedded. Consequently, string expressions were included in the previously created queries to identify library items that extract just the matching facet component. In some cases, namely mapping literature time period codes onto the Standard Subdivision Time Facet, translation tables are required.

5.3.6 Combining Sets

Population of each of the four facets with library items according to the three sources of Dewey codes entailed a series of different queries. These different queries are combined by a single Union query for each of the four facets, which also eliminates any duplicates within the query result sets. Only one query therefore is ultimately required to populate each facet.

Once this procedural method has been implemented and the preliminary queries created to culminate in a single union query for each facet, these queries could be applied to any collection that has been classified by Dewey. The new library item details would be imported into a database, together with the previously created queries which would be run against the new library collection. Once assembled, ultimately, this would allow each facet to be automatically populated by the single union query at any given time to create and update the facet population. The chosen four facets described in this doctoral research are case studies with a strong literary warrant, yet other libraries may prefer a different set of facets, tailored to their own collections and user needs. In this case, the same procedure would be followed, yet the actual Dewey query criteria would need to be adjusted accordingly, since different concepts have different Dewey notations.

5.4 Conclusion

This chapter has provided a conceptual overview of the components and procedure for deriving and populating facet views of the Dewey Decimal Classification Scheme for application in a library OPAC. As mentioned in the above section 5.3.5, these four facets are grouped into 'simple' (Form/Reference and Person) and 'complex' (Time and Place) respectively. The reasoning behind this grouping relates to their similar nature of derivation and population. The Form/Reference and Person facets are 'simple' in that they both have a flat structure (enabling additional title information to be utilised); the library items typically match the facet table, not requiring translation tables or string expressions to extract the relevant component. These 'simple' facets are also subject to a numerous false drops due to the lack of a uniform notation in Dewey, requiring exclusion heuristics to be developed and incorporated. In contrast, the Time and Place facets are hierarchical in structure (discouraging the use of title information), frequently requiring translation tables (i.e. time periods in literature by language) or string expressions to enable a match with the facet table. However, unlike the 'simple' facets, they are less prone to false drops, since a large proportion of the deconstruction is assisted by facet citation order (syntax), namely for the Time facet, rather than solely 'pattern matching' according to specific criteria. The following chapters 6 and 7 therefore illustrate and report the results of the devised method in practice, in the context of the four different facets; firstly discussing the 'simple' facets of Form/Reference and Person, followed by the more 'complex' facets of Time and Place.

CHAPTER 6

EMPIRICAL CASE STUDY 1: REFERENCE TYPE AND PERSON FACETS

6.1 Introduction

This chapter applies the method described in the previous Chapter 5 in an empirical case study of the derivation and population of two facets: Reference Type and Persons. These facets are discussed together since they can be regarded as 'simple'; their facet tables are flat in structure, unlike the remaining hierarchical facets of Time and Place, which will be discussed in the following Chapter 7. This chapter discusses each facet in turn, firstly explaining the rationale for their choice, followed by the creation of the Dewey facet tables, sources of the Dewey codes, identification and mapping of relevant library items to the Dewey facet tables and, finally, combination of the various result sets to 'populate' the facet.

6.2 The Reference Type Facet

6.2.1 Facet Identification

Reference sources are an important tool for users and librarians alike:

> Reference books are a means to the end of accurate, usually succinct responses to trivial or earth shattering queries. Unquestionably, from South Africa to Alaska, from 3000 BC to the present, there have been people with questions and reference works in which to find the answers. Organization, format and certainly audiences have changed drastically, yet the need for superficial or philosophical absolutes in response to questions remains the same.
>
> Katz (1997, p. 396)

The derivation and potential value of this facet is supported by the existence of internal library literature that directs users to 'key information/reference sources' in particular subject areas, produced by libraries university wide, including Huddersfield, and the importance of such reference sources to librarians for answering user enquiries. Moreover, although Dewey collates general reference works within the Generalities class, subject based sources are usefully located but simultaneously scattered by their relevant subject discipline and not searchable as a discrete collection. Some libraries, namely public libraries, have chosen to counter this in certain cases by providing physical

arrangements alterative to Dewey at the library shelves; for example, biographies are housed as a single collection rather than being located within their respective subject disciplines. This current research will enable such a user to view biographies as both a separate collection and with regard to their subject allegiance. The ability to identify reference works both within and across subject discipline may therefore prove extremely useful, offering users an alternative arrangement to that given at the library shelves.

6.2.2 Facet Creation

Although the categories within the Reference Type Facet are derived from the DDC, the coding scheme employs three letter codes, as follows:

ReferenceTypeID	ReferenceType
BIB	Bibliographies
BIO	Biographies
DEN	Dictionaries and Encyclopaedias
DIR	Directories

Table 6.1: Reference Type Coding Scheme in Facet Table

As shown in Table 6.1, this simple facet consists of just four categories with no hierarchical arrangement.

6.2.3 Sources of DDC Codes

The Dewey Decimal Classification Scheme enumerates general reference works such as dictionaries, encyclopaedias and bibliographies within the Divisions and Sections of the Main Class 000 Generalities. General *subject discipline based* reference works such as dictionaries are similarly enumerated for each Main Class using Standard Subdivision notations, for example:

103 *Dictionaries, encyclopedias, concordances of philosophy*
503 *Dictionaries, encyclopedias, concordances* (of Natural sciences and Mathematics)

Subject specific reference works are not generally enumerated but have their particular form notation synthesised using the Standard Subdivision (Table 1) notations, which can be appended to any number, therefore subject discipline, unless stated otherwise, for example:

T1--025 Directories of persons and organizations
T1--03 Dictionaries, encyclopedias, concordances
T1--092 Persons (Biographies)

Three sources were used for the identification of library items, namely:

1. Enumerated Schedule captions
2. Codes synthesized using the Standard Subdivisions
   - Main Class and Division Standard Subdivisions
3. Title keywords

In addition to these three main sources, other auxiliary tables (viz. *Table 4. Subdivisions of Individual Languages and Language Families)* were utilised for the Dictionaries and Encyclopedias category of the Reference Facet. These sources and subsequent result sets are now discussed in turn, culminating in their amalgamation using a single union query.

6.2.4 Identifying and Mapping Library Items to the Facet Table

6.2.4.1 Enumerated Schedule Captions

The Dewey Schedule captions were searched for the 4 categories of reference type, using the criteria shown in Table 6.2, to identify relevant codes. These codes were subsequently matched to library items in a series of queries and the relevant Reference Facet ID code assigned.

Reference Type	Query criteria	Number of Schedule DDC codes	Number of library items
Dictionaries & Encyclopedias	(Like dictiona* or like encyclop*) And not like *encylopedists*	44 (45 identified)	613
Directories	Like *directories*	9	109
Bibliographies	Like *bibliographies*	21	808
Biographies	Like *biograph*	7	64
Total			1594

Table 6.2: Reference Facet Query Criteria and Results for Enumerated Schedule Caption Search

Table 6.2 shows a total of 1594 reference type library items identified via the Schedule captions. Forty-four of the 45 codes identified for dictionaries and encyclopaedias were utilised, owing to exclusion of *019 Dictionary Catalogs*.

6.2.4.2 Standard Subdivisions

The Standard Subdivision codes and captions were firstly examined to identify Dewey codes, which were then incorporated into queries to retrieve library items. However, upon analysis of the results, it became apparent that there were a noticeable number of false drops caused by the re-use of codes for different concepts. For example:

Dewey Class Code	Caption
011.03	*Bibliographies of free materials
271.03	Teaching orders
321.03	Empires
342.03	Revision and amendment of the basic instruments of government
343.03	Law of public finance
344.03	*Social service
345.03	*Criminals (Offenders)
355.03	Military situation and policy
371.03	Community schools
547.03	*Oxy and hydroxy compounds
617.03	Rehabilitation
661.03	Metallic compounds
697.03	Central heating
937.03	Period of unification of Italy, ca. 500-264 B.C.

Table 6.3: Examples of Re-Use of 03 Standard Subdivision Code for Different Concepts

As shown by the few examples in Table 6.3, the -03 Standard Subdivision Code is re-used throughout the Schedules to represent alternative concepts. Consequently, the results were examined further to ascertain whether there were any notational patterns and to devise a means of excluding such false drops. Three exclusion heuristics were therefore identified for eliminating some false drops:

1. *Exclude if Standard Subdivision has subsequent facets*

Form divisions generally occur as the final facet in a citation order. Instances where this is not the case can be excluded by removing the closing truncation asterisk in the query criteria.

2. *Exclude if Standard Subdivision has a Schedule caption, i.e. is not a built number*

Although the code utilises a Standard Subdivision notation, in such cases it does not typically represent a Standard Subdivision concept, which instead may be signalled by multiple, as opposed to the regular single, zeros; however, due to the truncation immediately preceding the Standard Subdivision in the query criteria e.g. *.*03** and in the absence of exclusion criteria, both instances would be retrieved. For example:

661.03 Metallic compounds
661.001-661.009 Standard subdivisions

However, including a Dewey Schedule caption field within the query and the criteria Is Null will retrieve only those library items with built Standard Subdivision notations, excluding false drops such as 661.03 Metallic compounds enumerated in the Schedules. The few valid reference type Schedule codes which are enumerated in the Schedules will also be excluded but will have been previously retrieved by the Schedule captions query (See Section 6.2.4.1).

3. Exclude if Standard Subdivision is preceded by --09

The --09 facet indicator precedes an historical period or geographic place.

The following table details the query criteria for each of the four reference types, the number of library items before the application of exclusion heuristics to eliminate false drops, the type of heuristics applied and the number of library items after application of such heuristics:

Reference Type	Query criteria (including exclusion criteria)	No. Library Items Before Exclusion Heursitics	Exclusion heuristics applied	No. Library Items After Exclusion Heuristics
Dictionaries & Encyclopedias	Like "???.*03" And Not Like "*0903*" And Not Like "*0.903*" And Not Like "*09.*03*" And Not Like "*0.9*03*" And Not Like "0*9*03*" And Not Like "*09*103*" And Not Like "*0.9*103*" And Not Like "*.*09003*" And Not Like "616.[1-9]*03*" And Not Like "701.03*" And Not Like "720.103*" And Not Like "91[3-9].*03" And Not Like "941.03*"	3924	1-3	1026 (864 after exclusion of individual codes with 10+ library items)
Directories	(Like "???.*025*" Or Like "???.*0294*") And Not Like "*09025*" And Not Like "*0.9025*" And Not Like "*09.*025*" And Not Like "*0.9*025*" And Not Like "*.*09*025*" And Not Like "*.*02509*" And Not Like "016.025*" And Not Like "34[2-9].*025*" And Not Like "616.[1-9]*025*" And Not Like "618.92*025*" And Not Like "808.025" And Not Like "9[3-9]?.*025*"	431	2-3	196 (149 after exclusion of individual codes with 10+ library items)
Bibliographies	Like "???.*016" And Not Like "599.016"	382	1-2	305
Biographies	Like "???.*092[2-4]*"	1676	2	1676

Table 6.4: Reference Facet Query Criteria for Standard Subdivision Search and Effect of Exclusion Heuristics

It can be seen from Table 6.4 that in the case of Dictionaries and Encyclopaedias all three exclusion heuristics were applied, reducing the initial set by 2898 library items (74%). If we examine these dictionary and encyclopaedia query results in terms of unique Dewey codes via a Select Distinct query, thus disregarding the number of library items, then the initial result set with no exclusion criteria amounts to 1007 distinct DDC codes, reduced by 453 codes (45%) to 554 after application of heuristic 1, reduced by a further 64 codes (6%) to 490 with application of heuristic 2, and finally another 86 codes (9%) to 404. Application of the three heuristics therefore reduced the initial distinct result set by 603 codes, 60% (i.e. % of false drops) in total. Table 6.5 below shows example false drops,

which would be eliminated by heuristics 1 - 3, namely suppressing library items with digits following 03 (e.g. those coloured red), 09 notations (e.g. LibraryItemID 717, coloured green) and those having Schedule captions, indicating irregular re-use of Standard Subdivision codes. However, it is noted that since all the codes in the table have captions (some also additional digits and one an 09 component) heuristic 2 alone would in fact exclude all of these.

LibraryItemID	DeweyCode	Title	Schedule Caption
717	709.032	Baroque and Rococo	*17th century, 1600-1699
1116	966.903	Sir George Goldie and the making of Nigeria	Period as a British colony, 1886-1960
1527	371.103	Parent power : securing the best schooling for your child	Teacher-parent conferences
1533	951.032	China and the West, 1858-1861 : the origins of the Tsungli Yamen	1644-1795
1543	378.103	Building a learning culture : a report of a conference on personal and career development plans in higher education and employment, [held at Unilever Research Colworth Laboratory, near Bedford in March 1993]	Community relations
1651	658.403	The design of the management information system	Decision making and information management
1667	658.4034	Operations research : an introduction to modern applications	Operations research

Table 6.5: Example Dictionary/Encyclopaedia 'False Drops' Eliminated by Exclusion Heuristics

Despite applying these three heuristics several false drops within the dictionary and encyclopaedia set were still apparent, namely approximately 200 library items having 30 distinct codes. However, the majority of these 30 distinct false drop codes are assigned to only one or two library items. There seem to be two causes of such false drops even after the three heuristics have been applied.

The first is that many of these codes are expressed as ranges in the Dewey Schedules and thus the join query on library item codes and Dewey Schedule codes was unsuccessful in recognising and excluding these library item codes as non-dictionary enumerated

Schedule codes (heuristic 2). For the present query to be successful, these ranges would need to be broken down which, manually, would take considerable time.

The second cause is where a number is constructed from a Schedule table with the instruction "Add as instructed..." or from part of another Schedule codes under the instruction "Add to base number..". It was decided that those with a more significant number of library items, i.e. 10 plus, should be incorporated into the query exclusion criteria, as shown in Table 6.3. This action provided a final set of 864 library items.

It should be noted, however, that the above codes are particular to the University of Huddersfield library collection and thus may be of less use as exclusion criteria for other library collections.

As shown in the query criteria in Table 6.4, two Standard Subdivision codes were utilised to identify directories, namely:

--025 Directories of persons and organizations
--0294 Trade catalogs and directories

False drops were removed from the initial result set of 431 library items by applying heuristics 2 (more than halving the set to 221) and 3, providing a small reduction to 196 library items. As previously, false drops remained even after such exclusions heuristics had been applied. Since no common pattern distinguished these false drops, individual codes having more than 10 library items were again incorporated into the query exclusion criteria, as detailed in Table 6.4. The resultant set comprised 149 directory-related library items.

False drops in the bibliography result set were similarly examined and excluded by way of heuristics 1 and 2, representing a reduction from 382 to 325. Only one individual Dewey code was further excluded: 599.016 (599.01-08 Mammalia General principles) from the 19th edition of Dewey and no longer enumerated in the current edition. The final set contained 305 library items.

Finally, biographies were identified, utilising the following Standard Subdivision codes:

--092 Persons
--0922 Collected persons treatment
--0923 Collected persons treatment of members of specific racial, ethnic, national groups
--924 Individual biography (19th edition of Dewey)

Although heuristic 2, the addition of captions to eliminate enumerated Schedule codes, was applied, in the event this was not required since all the library item codes were synthesised, evidenced by the fact that the result set remained at 1676 library items.

Therefore, of the form division codes described above, it appears that biography/persons are the only set of Standard Subdivision notations which are not re-used to represent other concepts in the Schedules. Such exclusiveness is extremely beneficial, obviating the need to eliminate false drops.

6.2.4.3 Main Class and Division Standard Subdivisions

Main Class Standard Subdivision codes, e.g. 103 Philosophy and Psychology dictionaries/encyclopedias are already enumerated and hence identified by the previous Schedule caption queries (Section 6.2.4.1).

Queries were constructed to identify the Division Standard Subdivisions for each of the four categories in the Reference Facet, in which the zero section digit is dropped before the Standard Subdivision code is appended, e.g. Psychology dictionaries 150 - 0 + 03 = 150.3.

Reference Type	Query criteria	No. Library Items
Dictionaries & Encyclopedias	Like "1?0.3" Or Like "[4-6]?0.3" Or Like "8?0.3" Or Like "[2-3]?0.3" Or Like "7?0.3"	221
Directories	(Like "1?0.25*" Or Like "[4-6]?0.25*" Or Like "8?0.25*" Or Like "[2-3]?0.25*" Or Like "7?0.25*") And Not Like "630.25*"	24
Bibliographies	(Like "1?0.16" Or Like "[4-6]?0.16" Or Like "8?0.16" Or Like "[2-3]?0.16" Or Like "7?0.16") And Not Like "530.16" And Not Like "690.16"	4
Biographies	Like "[1-8]09.2[2-4]*" Or Like "1?0.92[2-4]*" Or Like "[4-6]?0.92[2-4]*" Or Like "8?0.92[2-4]*" Or Like "[2-3]?0.92[2-4]*" Or Like "7?0.92[2-4]*"	1920

Table 6.6: Reference Facet Query Criteria for Division Standard Subdivisions

As in the case of the regular Standard Subdivisions, heuristics 1 and 2 were applied to exclude false drops. As can be seen in Table 6.6, individual Dewey codes within ranges, built 'Add to base' numbers and codes from earlier editions of the DDC were excluded for directories and bibliographies, namely:

630.25 (within range 630.21-630.29 Scientific principles)
530.16 Measurement theory (19th edition)
690.16 (690.1 Structural elements Add to base number 690.1 the numbers following 721 in 721.1-721.8, e.g., auxiliary roof structures 690.15)

6.2.4.4 Other Auxiliary Tables

In the case of dictionaries and encyclopedias, Table 4 is applied under an 'Add to base number' instruction to codes *420-490 Specific languages,* with the extensions:

--03 Encyclopedias and concordances
--3 Dictionaries of the standard form of the language
--31 Specialized dictionaries
--32- --39 Bilingual dictionaries

Therefore, initially a query was constructed with the criteria 4[2-9]?.*3"; however, upon examination of the result set and even after application of heuristic 2 (i.e. caption 'Is Null' to retrieve only built codes), the result set was still mainly non-dictionary or contained library item codes which would have been retrieved by the previous regular Standard Subdivisions query (Section 6.2.4.2). The exceptions were *429.3 Old English (Anglo-Saxon) - Dictionaries* and *442.3 Etymology of standard French - Dictionaries* and certain 49? *Other languages* codes whereby, unlike languages allocated their own Division language code and enumerated dictionary/encyclopedia Section code and hence retrieved by the Schedule Caption query (for example, 47**3** *Dictionaries of classical Latin)* the notation is built. The query was therefore re-written as follows:

(Like "49?.*3" Or Like "429.3*" Or Like "442.3*") And Not Like "49?.*03"

This query resulted in just 17 library items. 49?.*03 codes were excluded since they had been extracted by the previous Standard Subdivision query.

6.2.4.5 Title keywords

Title keywords were utilised to populate the Reference Facet as a supplement to the Dewey classification, providing the following results:

Reference Type	Query criteria	No. Library Items
Dictionaries & Encyclopedias	Like "*dictiona*" Or Like "*encyclop*"	1842
Directories	Like "*directories*" Or Like "*directory*"	497
Bibliographies	Like "*bibliograph*"	693
Biographies	Like "*biograph*"	737

Table 6.7: Reference Facet Query Criteria for Title Keyword Search

6.2.5 Combining Sets

The queries detailed in sections 6.2.4.1-6.2.4.5 were combined using a series of union queries for each of the four categories in the Reference Facet. These four union queries were then subsequently amalgamated into a single union query to give a total number of 'Reference Type' library items:

Reference Type	No. Library Items
Dictionaries & Encyclopedias	3556
Directories	778
Bibliographies	1809
Biographies	4396
Total	10,539

Table 6.8: Cumulative Totals for Reference Facet

6.3 The Person Facet

6.3.1 Facet Identification

The impetus behind deriving a Person Facet followed from a personal discussion with Julianne Beall (Assistant Editor of the DDC), who highlighted the problems faced by students of the interdisciplinary subject women's studies, whereby the literature is considerably scattered across the classification scheme. Such scatter was resolved by one particular further education library where the current author worked, by housing the women's studies literature in a separate sequence. However, derivation of a person facet for presentation at the OPAC interface precludes the need to create a separate collection, allowing a user to view and search both women's studies with respect to its disciplinary treatment (the DDC shelf arrangement) and as a topical entity in itself.

Two further categories comprise the Person Facet, namely Men and Children. It is recognised that this facet does not strictly adhere to the principles of faceted classification, in that more than one characteristic is applied: by gender (Women and Men) and by age (adults - implicit in women/men - and children). The choice of these categories was a pragmatic decision, based on an assumption that this would be the most explicit and helpful grouping for users, as opposed to 'male/female', for example.

6.3.2 Facet Creation

Like the Reference Facet, although the categories are derived from the DDC, the facet coding scheme comprises a series of three letter codes:

PersonTypeID	PersonType
KID	Children
MEN	Men
WOM	Women

Table 6.9: Person Type Coding Scheme in Facet Table

6.3.3 Sources of DDC Codes

Four sources were used to identify codes to populate each of the categories in the Person Facet:

1. Enumerated Schedule Captions
2. Codes Synthesized using the Standard Subdivisions
    - Main Class and Division Standard Subdivisions
3. Other Auxiliary Tables (3B, 3C and 7)
4. Title keywords

6.3.4 Identifying and Mapping Library Items to the Facet Table

6.3.4.1 Enumerated Schedule Captions

The Schedule Captions were searched to identify Dewey codes representing each of the three categories in the Person Facet. These codes were then matched against the library collection to identify library items.

Person Type	Query criteria	Number of Schedule DDC codes	Number of library items
Women	(Like "*women*" Or Like "*female*" Or Like "*femini*") And Not Like "* men*" And Not Like "men*"	97	1482
Men	(Like "* men *" Or Like "men *" Or Like "men" Or Like "* men" Or Like "men's *" Or Like "man *" Or Like "* man *" Or Like "man" Or Like "* man" Or Like "* male *" Or Like "male *" Or Like "male" Or Like "* male" Or Like "*masculini*") And Not Like "*women*" And Not Like "*female*"	61	90
Children	Like "*child*" Or Like "*infants*" Or Like "*infant *"	108	1589
Total			3161

Table 6.10: Person Facet Query Criteria and Results for Enumerated Schedule Caption Search

Table 6.10 shows a total of 3161 library items identified using the Schedule captions. A further query was constructed to identify Schedule captions that refer both to men and

women, retrieving 4 DDC codes; however, as signified by the bracketed discontinued notations, only 2 of these are still in use:

Dewey Class Code	Caption
646[.3081-646.3082]	Clothing for men and women
267.1	Associations for religious work for both men and women
305.3	Men and women
920[.0081-920.0082]	Men and women

Table 6.11: Men and Women in Schedule Captions

The above codes retrieved 152 library items; however, it was subsequently decided against having a separate query in favour of incorporating these codes into both the Women and Men Schedule caption queries.

6.3.4.2 Standard Subdivisions

Examination of the Standard Subdivisions identified candidate 'Person' DDC codes, as follows:

T1--082 Women
T1--08352 Women 12-20
T1--08422 Young Women
T1--086643 Lesbians

T1--081 Men
T1--08351 Males twelve to twenty
T1--08421 Young men
T1--086642 Gay men

T1--0832 Infants
T1--0833 Children three to five
T1--0834 Children six to eleven
T1--086/945 Abandoned children, abused children, children born out of wedlock, orphans

These codes were incorporated into three queries to identify library items for each category in the Person Facet:

Person Type	Query criteria (including exclusion criteria)	No. Library Items Before Exclusion Heursitics	Exclusion heuristics applied	No. Library Items After Exclusion Heuristics
Women	(Like "???.*082" Or Like "???.*08352*" Or Like "???.*08422" Or Like "???.*086643") And Not Like "*.*09*082*" And Not Like "??0.9*082*" And Not Like "?09.*082*" And Not Like "016.*082*" And Not Like "354.*082*"	417	1-3	126 (95 after exclusion of individual codes with 10+ library items)
Men	(Like "???.*081" Or Like "???.*08351*" Or Like "???.*08421" Or Like "???.*086642") And Not Like "*.*09*081*" And Not Like "??0.9*081*" And Not Like "?09.*081*" And Not Like "94[1-8].08*" And Not Like "94[1-8].*08*" And Not Like "???.*94[1-8]08*" And Not Like "???.94[1-8]*08*" And Not Like "27[4-9]*" And Not Like "283.[4-9]*" And Not Like "354.420081*" And Not Like "530.081*" And Not Like "016.7808*"	387	1-3	35 (8 after exclusion of individual codes with 10+ library items)
Children	Like "???.*083[2-4]*" Or Like "???.*086945*"	10	2	8

Table 6.12: Person Facet Query Criteria for Standard Subdivision Search and Effect of Exclusion Heuristics

As in the Standard Subdivision Reference Facet query, false drop exclusion heuristics were applied. Enumerated Schedule codes were eliminated from the outset (heuristic 2), for example:

LibraryItemID	DeweyCode	Title	DeweyCaption
2514	944.082	Charles de Gaulle : futurist of the nation	Period of Fourth Republic, 1945-1958
2989	941.082	The Edwardian age	1901-1999
3542	343.082	Food labelling : a guide to the statutory requirements	Advertising and labeling
5353	946.082	Espaäna en crisis : evoluciâon y decadencia del râegimen de Franco	Period of Francisco Franco, 1939-1975
5457	342.082	Slamming the door : the administration of immigration control	Entrance to and exit from national domain

Table 6.13: Example Person Facet 'False Drops' Eliminated by Exclusion Heuristics

Further heuristics were applied recursively upon examination of the result sets. The difference between the number of library items before and after the exclusion heuristics are applied, as detailed in Table 6.13, indicates a substantial number of false drops (Women 70% and Men 91%) and hence the re-use of Standard Subdivision codes, hindering the deconstruction process.

In addition, a limited number of individual false-drop DDC codes, typically within notational ranges or from earlier DDC editions, were excluded for the Women and Men categories. The result set for 'Men' also contained another source of false drops: European history codes 941-948, which occur both within the main class 900 and also embedded within other main classes, whereby 08 signifies an historical period, for example:

LibraryItemID	DeweyCode	Title
2468	324.941505081	Parliamentary election results in Ireland, 1801-1922
162525	324.942817081	Nineteenth century Bradford elections
88515	376.941081	The Education papers : women's quest for equality in Britain, 1850-1912
110321	769.49941081	A social history of Britain in postcards 1870-1930
101706	942.081	A social history of England 1851-1990
92687	942.67081	The Great English Earthquake
90628	945.081	Fascism in Italy : society and culture, 1922-1945

Table 6.14: Example Person Facet 'False Drops'

6.3.4.3 Main Class and Division Standard Subdivisions

The Main Class and Divisions (first and second hierarchical levels) were examined to ascertain those codes having a '*...with respect to kinds of persons*' caption. The previously identified Standard Subdivision codes were then appended to these (with terminal zeros being dropped, as previously) to formulate the query criteria, however, with very disappointing results:

Person Type	Query criteria	No. Library Items
Women	Like "108.2" Or Like "408.2" Or Like "604.82" Or Like "908.2" Or Like "108.352" Or Like "408.352" Or Like "604.8352" Or Like "908.352" Or Like "108.422" Or Like "408.422" Or Like "604.8422" Or Like "908.422" Or Like "108.6643" Or Like "408.6643" Or Like "604.86643" Or Like "908.6643" Or Like "??0.82" Or Like "??0.8352" Or Like "??0.8422" Or Like "??0.86643"	12
Men	Like "108.1" Or Like "408.1" Or Like "908.1" Or Like "108.351" Or Like "408.351" Or Like "908.351" Or Like "108.421" Or Like "408.421" Or Like "908.421" Or Like "108.6642" Or Like "408.6642" Or Like "908.6642" Or Like "?[0-8]0.81" Or Like "??0.8351" Or Like "??0.8421" Or Like "??0.86642"	0
Children	Like "108.3[2-4]*" Or Like "408.3[2-4]*" Or Like "908.3[2-4]*" Or Like "108.6945*" Or Like "408.6945*" Or Like "908.6945*" Or Like "?[0-8]0.83[2-4]*" Or Like "??0.86945*"	0

Table 6.15: Person Facet Query Criteria for Division Standard Subdivisions

6.3.4.4 Other Auxiliary Tables

In addition to the Standard Subdivisions, Dewey codes relating to women, men and children appear in the other auxiliary tables, added under instruction in the Schedules:

Person Type	*Table*	Dewey Code	Caption
Women	3B	--080352042	Women--literature--specific literatures
		--08352042	Women--literature--specific literatures--history and criticism
		--099287	Women authors (Literature)--specific literatures
	3C	--352042	Women--arts
		--9287	Women
	7	--042	Females
Men	3B	--080352041	Men--literature--specific literatures
		--09352041	Men--literature--specific literatures--history and criticism
	3C	--352041	Men--arts
		--9286	Men
	7	--041	Males
Children	3B	--080352054	Children--literature--specific literatures
		--09352054	Children--literature--specific literatures--history and criticism
	3C	--352054	Children--arts
		--9282	Children
	7	--0542	Infants
		--0543	Children three to five
		--0544	Children six to eleven

Table 6.16: Person Codes in Other Auxiliary Tables

Queries were constructed to identify library items with Table 3 codes, applied to the 800 *Literature (Belles-lettres) and rhetoric* and 700 *The arts Fine and decorative arts* (viz. Table 3C) classes, as follows:

Person Type	***Criteria***	**Number of Library Items**
Women	Like "8??.*080352042*" Or Like "8??.*08352042*" Or Like "8??.*099287*" Or Like "8??.*352042*" Or Like "8??.*9287*" Or Like "700.452042*" Or Like "791.*52042*"	166
Men	Like "8??.*080352041*" Or Like "8??.*09352041*" Or Like "8??.*352041*" Or Like "8??.*9286*" Or Like "700.452041*" Or Like "791.*52041*"	1
Children	Like "8??.*080352054*" Or Like "8??.*08352054*" Or Like "8??.*352054*" Or Like "8??.*9282*" Or Like "700.452054*" Or Like "791.*52054*"	4

Table 6.17: Person Facet Query Criteria for Table 3 Codes

Heuristic 2 (exclusion of codes with captions to retrieve only built numbers) was also applied as criteria. The criteria for the Table 3C codes, appended to the Dewey codes 700.4 and 791.4 was constructed based on instructions in the Schedules, which required the initial digit in the Table 3C code (see Table 6.16 for examples) to be dropped, for example:

> *700.42-700.48 Arts dealing with specific themes and subjects*
> Add to base number 700.4 the numbers following T3C--3 in notation T3C--32-T3C--38 from Table 3-C, e.g., historical themes in the arts 700.458

As can be seen from Table 6.17, with the exception of 'Women', the result set for this particular library collection is rather disappointing, with minimal retrievals.

Library item DDC codes containing Table 7 notations were identified via a two-stage process. Firstly a query was constructed to search the Schedule notes field for add instructions using "*Table 7*", retrieving 14 codes and captions.

Dewey Class Code	Note
013.03-013.87	Add to base number 013 notation T7--03-T7--87 from Table 7, e.g., women authors 013.042
013.89	Add to base number 013.89 the numbers following T7--9 in notation T7--91-T7--99 from Table 7, e.g., archaeologists 013.893
028.53	Add to base number 028.53 the numbers following T7--05 in notation T7--054-T7--055 from Table 7, e.g., reading and use of other information media by young people twelve to twenty 028.535
133.594	Add to base number 133.594 the numbers following T7--29 in notation T7--292-T7--299 from Table 7, e.g., Hindu astrology 133.59445
174.9	Add to base number 174.9 notation T7--09-T7--99 from Table 7, e.g., ethics of genetic engineering 174.957 ; however,
303.388	Add to base number 303.388 notation T7--04-T7--99 from Table 7, e.g., opinions of dentists 303.3886176 ; then add 0* and to the result add notation T2--1-T2--9 from Table 2, e.g., the opinions of dentists in France 303.3886176044
305.43	Add to base number 305.43 notation T7--09-T7--99 from Table 7, e.g., female physicians 305.4361
305.6	Add to base number 305.6 the numbers following T7--2 in notation T7--21-T7--29 from Table 7, e.g., Christian Scientists 305.685 ; then add 0* and to the result add notation T2--1-T2--9 from Table 2, e.g., Christian Scientists in France 305.685044
305.904	Add to base number 305.904 the numbers following T7--04 in notation T7--043-T7--046 from Table 7, e.g., grandchildren 305.90442
305.908	Add to base number 305.908 the numbers following T7--08 in notation T7--081-T7--082 from Table 7, e.g., gifted persons 305.90829
305.909-305.999	Add to base number 305.9 notation T7--09-T7--99 from Table 7, e.g., persons occupied with religion 305.92, postal workers 305.9383
332.02403-332.02499	Add to base number 332.024 notation T7--03-T7--99 from Table 7, e.g., personal finance for single people 332.0240652
390.4	Add to base number 390.4 notation T7--09-T7--99 from Table 7, e.g., customs of lawyers 390.4344 ; however,
704.04-704.87	Add to base number 704 notation T7--04-T7--87 from Table 7, e.g., women as artists 704.042 ; however,

Table 6.18: Schedule Codes With Add Notes for Table 7 Codes

The above codes were then examined to identify those where a Person Type extension was applicable; the identified codes were incorporated into a second query for each of the three Person Type categories, as follows:

Person Type	*Criteria*	Number of Library Items
Women	Like "013.042*" Or Like "303.388042*" Or Like "332.024042*" Or Like "704.042*"	3
Men	Like "013.041*" Or Like "303.388041*" Or Like "332.024041*" Or Like "704.041*"	0
Children	Like "013.054[2-4]*" Or Like "028.53054[2-4]*" Or Like "303.388054[2-4]*" Or Like "332.024054[2-4]*" Or Like "704.054[2-4]*""791.*52054*"	0

Table 6.19: Person Facet Query Criteria for Notations Incorporating Table 7 Codes

However, as Table 6.19 demonstrates, again the size of the result sets was disappointing.

6.3.4.5 Title keywords

Three queries were constructed with the following criteria and number of library items:

Person Type	*Criteria*	Number of Library Items
Women	Like "*wom?n*" Or Like "*female*" Or Like "*femini*"	3017
Men	Like "* man *" Or Like "* men *" Or Like "* men's *" Or Like "*masculin*"	769
Children	Like "*child*" Or Like "*infant*" Or Like "*baby*" Or Like "*babies*"	3820

Table 6.20: Person Facet Query Criteria for Title Keyword Search

As can be seen from Table 6.20, these queries retrieved quite substantial result sets. However, there were a number of probable false drops in the 'Men' set. For example, the fact that a title may mention 'man', it is not a reliable indicator that a library item is about males. The most obvious example is 'man' used in a generic sense to represent 'mankind', for example:

An essay on man : an introduction to a philosophy of human culture
The expression of the emotions in man and animals
The dirty man of Europe : the great British pollution scandal
The molds and man : an introduction to the fungi
The study of man : an introduction to human biology
The arts and man : a world view of the role and functions of the arts in society

Another example is when the item about or from the perspective of individual, who happens to be male. As can be seen below, typical examples include when both an individual's professional (e.g. 'the scientist') and personal life (e.g. 'the man') are discussed:

Alan McGee-the man who discovered Oasis
Trouble in paradise: Clive Anderson, our man in... Havana
Second chance : the true story of a man who 'died' and lived to describe the experience
Neill and Summerhill : a man and his work
William Henry Bragg, 1862-1942 : man and scientist
Henry Dreyfuss, industrial designer : the man in the brown suit
Gustav Holst : the man and his music

Other false drops include proper nouns such as the Isle of Man and other languages such as the German example below.

Andy Warhol, 1928-1987 : works from the collections of Josâe Mugrabi and an Isle of Man company
The Isle of Man : a social, cultural and political history

So schreibt man Briefe besser

Thus, in terms of accuracy (precision), in this instance, classification seems more superior to cryptic or misleading titles, as made explicit by Dewey's manual notes (OCLC, 1998):

T1--081 and T1--082, T1--08351, T1--08352, T1--08421, T1--08422
Subdivisions for men and women should be used only for works explicitly emphasizing the sex of the people treated. For example, do not use 363.37081 for men as a group in respect to fire fighting unless the work makes clear that firemen are being contrasted with firewomen, or 364.3608351 for juvenile delinquents (a term often implying young men under eighteen) unless male delinquents are being contrasted to female delinquents.

Another issue was that the 'Women' and 'Men' queries also retrieved titles containing both keywords, i.e. not exclusively men or women. Initially, it was investigated whether to create a joint Men and Women category (see also Table 6.11 for Schedule captions

referring to both men and women). A query was created to retrieve both these title keywords, with the criteria:

> Like "* man * *woman*" Or Like "* men * women*" Or Like "* men's * women's*" Or Like "*masculin* femin*" Or Like "* male * female*" Or (Like "*woman* * man *" Or Like "* women* * men *" Or Like "women's* * men's *" Or Like "*femin* mascluin*" Or Like "*female* male *")

This query identified 77 library items and was then matched to the previous 'men' and 'women' title keyword queries by way of two inner join queries with the 'Is Null' criteria in the title field. These queries excluded items with *both* men and women in the title (i.e. where there was no match) thus leaving remaining sets of either 'women' or 'men' only.

Despite a successful attempt to resolve this issue, in the event, it was decided to return to the original query since it was thought that it would be more helpful to the user. Therefore, the result sets of the men and women title queries will contain library item titles with just women or men, yet also those referring to both 'women' and 'men' simultaneously.

6.3.5 Combining Sets

The queries detailed in sections 6.3.4.1-6.3.4.5 were combined by way of three union queries for each Person Type category, which were then brought together into a single union query, with the following results:

Person Type	No. Library Items
Women	4774
Men	867
Children	5420
Total	11,061

Table 6.21: Cumulative Totals for Person Facet

6.4 Conclusion

This chapter has described the derivation and population of the Reference Type and Person Facets, according to the method described in the previous Chapter 5. It was described how the creation of such facets could provide alternative arrangements to the predominant model for physical library collections, which frequently scatters such types of material across the classification scheme, library collection and even the building itself. Having established a clear rationale for each facet, facet meta data was created and sources of Dewey codes successfully identified. The sources were identical for both facets (Enumerated Schedule Captions, Synthesised Standard Subdivision Codes and Title Keywords), apart from additional auxiliary table sources used in the Persons Facet. Based on these sources, a series of SQL queries was then devised to identify library items. These queries successfully identified relevant library items; however, a number of problems were encountered in the process. The main difficulty, with the exception of the 'biography' category in the Reference Facet, was the re-use of Standard Subdivision Dewey codes, also highlighted in Chapter 4, which resulted in several 'false drops' (non-relevant library items). As a result, three exclusion heuristics were devised to help counteract such problems, achieving encouraging results. Despite the application of these heuristics, certain false-drop codes still remained, namely those represented by ranges in the Schedules or those synthesised using a Schedule Table. However, it was found that these non-excluded 'false drops' only represented a small proportion of the result set; hence, it was decided to exclude manually only those codes having more than 10 library items. This manual intervention is clearly not ideal, but it will remain a problem until the DDC moves towards a more uniform notation. The following chapter further investigates and corroborates these findings by the derivation and population of two complex facets: Time and Place.

CHAPTER 7

EMPIRICAL CASE STUDY 2: TIME AND PLACE FACETS

7.1 Introduction

This chapter further exemplifies the method described in Chapter 5 by examining the second set of case studies: Time and Place. In contrast to the previous case study, the derivation and population of these facets can be regarded as 'complex' in that the facet table structure, originating from the DDC, is hierarchical, requiring many-to-many mapping and also extraction of the embedded facet component in order to match and populate the facet. These two facets are populated solely by DDC criteria; for reasons explained in Chapter 5, library item titles are not used. Moreover, in addition to the previously utilised pattern-matching queries, queries based on positional criteria (syntax) were also employed. This chapter discusses each facet in turn, firstly explaining the rationale for their choice, followed by the creation of the Dewey facet tables, sources of the Dewey codes, identification and mapping of relevant library items to the facet tables and, finally, combination of the various result sets to 'populate' the facet.

7.2 The Time Facet

7.2.1 Facet Identification

Designated as a fundamental facet by Ranganathan, an historical viewpoint is central to study in many subject disciplines. This is substantiated by strong literary warrant, with numerous library item titles published such as:

Rejection and emancipation, writing in German-speaking Switzerland 1945-1991
The poet reclining, selected poems 1962-1980
The historical mode - fashion and art in the 1980s

Overwhelmingly, (excluding the subject of history) the time period facet takes low priority in the DDC facet citation order and is consequently scattered across the notation and the library building. Creation and collation of library items by a time facet enables a user to view everything classified as being about a particular time

period in general (the top level hierarchy) and also more specifically by selecting a particular time period within the hierarchical structure of the Time Facet, for example the 1960s. The user can also then revert to a distribution view of these time-related library items according to their original subject discipline-based shelf arrangement.

7.2.2 Facet Creation

The structure for the Time Facet is taken from the 0901-0905 century notations found in the Standard Subdivisions (Table 1) of the DDC, namely:

--0901 To 499 AD
--0902 500-1499
--0903 Modern Period, 1500-
--0904 20th Century
--0905 21st Century

The above shows the top level of the hierarchy, from which a user can then select a category to reveal more specific time periods. This 0901-0905 hierarchy was reproduced as a Table in our Access database, and will subsequently form the 'Time Facet view' in the View-based OPAC.

7.2.3 Sources of DDC Codes

In the Dewey Decimal Classification Scheme, the category of time is incorporated in several ways. For subjects such as history, where time is naturally the subject of study, the notations and captions are included in the Schedules via an enumerated notation or instructions to re-use Schedule code components. In such Dewey classes, time is represented by historical period, particular to a specific country, rather than a simple organisation by century, for example:

941.013-941.019 Pre-Anglo-Saxon period through reign of Saxon kings, 410-1066

It was envisaged that if such time-related Schedule codes were identified and extracted then these historical periods could be mapped against the general century time classification found in the Standard Subdivisions (Table 1), i.e. each historical period code also receives an additional general time notation which, once incorporated into the View-based system, will provide a general 'Time Facet' view. This will enable a user to either view library items about the history of specific countries according to the historical periods of each, or take a more holistic view of

library items by a chosen century. This prospect may benefit users who are examining world events across both time and place.

Similarly, for individual literatures of specific languages (800s) particular historical periods are important but in this case are added using language-specific period tables included at each Division in the 800s, the most faceted (viz. language, literary form, literary period) of the Dewey Main Schedule Classes. Although, in the case of general literary works, where no particular literary period other than century applies, numbers are built according to instructions to add from the century classification found in the 0901-0905 Standard Subdivisions.

As with all Standard Subdivision codes, the 0901-0905 general time classification can be added (unless otherwise instructed) with or without instruction to any other notation in the Schedules. In summary, then, three main sources were used to extract time-related library items and Dewey codes:

1. Enumerated Schedule captions
2. Codes synthesised using the 0901-0905 Standard Subdivisions
   - Main Class and Division Standard Subdivisions.
3. Synthesised codes from 800 literature class
   - Literature Generalia
   - Literature by language

The following sections describe how the above three sources were utilised to identify time-related library items, and how their DDC codes were deconstructed to extract (and in certain cases translate) the 'time' code component for mapping onto the appropriate time-period category in the Time Facet Table. Each discussion of the identification and mapping process according to the three main sources concludes with a description as to how the subset query results are combined with a union query. The final section details the combination of these interim union queries into a single union query to populate the Time Facet.

7.2.4 Identifying and Mapping Library Items to Facet Table

7.2.4.1 Schedule Captions

'Time' is difficult to locate in the Dewey Schedule captions since it is a rather abstract concept, occurring under various guises (i.e. there is no single caption term and hence search term to represent 'time'). The Dewey Schedule captions were examined to ascertain the different formats under which 'time' was incorporated. This enabled construction of a query and a search of the DDC Schedule captions:

*Like "*historical periods*" Or Like "*century*" Or Like "*centuries*" Or Like "*####*"*

The above query resulted in 2750 DDC codes and captions; however, a cursory examination identified numerous 'false drops' due to inclusion of the hash characters, to extract numerical values, (*####*) in the query criteria. Although this numerical component retrieved dates, it also identified DDC relocations in which the former number is enclosed in square brackets after the caption e.g. Education [formerly 370.19]. These relocations appear in this standard format; thus, *'And Not Like "*formerly*"'* was incorporated into the query criteria to remove these 206 false drops, leaving 2544 time-related codes and captions. The only exception contained both 'formerly' *and* a valid time period; however, there were no items in the library collection with this code:

949.5072 Period of monarchy, 1833-1924 [formerly 949.506]

The query retrieving these 2544 time codes was then matched against the library collection (Library Item table) in an Inner Join query on an 'Assignable Dewey code' field. This join query retrieved 8506 library items and, following a change from a SELECT to SELECT DISTINCT query which eliminated duplicate library item codes, 628 unique time-related library item Schedule codes.

An exact match is prerequisite for the join query to succeed; hence library items extended beyond the Schedule notation (i.e. built numbers) would be excluded, if remaining in their current 'built' form, as they would not match the Schedule caption codes, despite being time-related. To enable their inclusion, an 'Assignable Dewey' code field had previously been created within the library item table, containing library

item codes which were trimmed, where applicable, of the built part of their code to leave the basic Schedule notation, for example:

LibraryItemID	Title	PublicationYear	DeweyCode	AssignableDeweyCode
101719	The deserted village : the diary of an Oxfordshire rector James Newton of Nuneham Courtenay 1736-86	1992	941.07092	941.07

Table 7.1: Example of a Library Item Title with Built and Schedule Dewey Codes

The above example shows how a library item with the Dewey code 941.07092 is reduced to its Schedule code of *941.07 Period of House of Hanover, 1714-1837*, omitting the built Standard Subdivision notation *T1--092 Persons.*

Having now extracted a set of historical period Schedule codes and library items, the next step was to map these codes onto the general time classification in Table 1. Owing to the lack of commonality amongst these Schedule codes and the thought and decision-making required in the mapping process, it was considered best to do this manually. The large number of Dewey codes, for cost-benefit reasons, precluded mapping the entire 628 set so priority was given to codes with a larger number of library items. To establish a subset, the previous SELECT DISTINCT join query was adapted to COUNT the number of library items assigned to each Dewey code. Examining the results, it was decided to take 10 library items as the cut off point; the criteria >= 10 was incorporated and the query changed to a Make-table query, extracting a more manageable set of 169 codes.

The library item >= 10 count, which created this subset, was then summed to ascertain what proportion of the entire set of time-related Schedule library items was represented by the subset:

	TIME SCHEDULE CODES	LIBRARY ITEMS
FULL SET	628	8506
SUBSET >=10	169	7670
SUBSET % OF FULL SET	27	90

Table 7.2: Time-related Schedule Codes and Number and Percentage of Library Items

As can be seen from Table 7.2, only just over a quarter (27%) of the time-related Schedule codes constitute almost the entire set of library items (90%), hence justifying 10 library items as being an economical and effective cut-off point.

This subset table of 169 historical period codes was then matched, manually, to the century time classification (our time facet) found in the Standard Subdivisions. The original idea was to match and assign additional Dewey codes solely based on the classificatory codes and captions, e.g.:

DeweyClassCode	Caption	FacetCodeID	TimeFacetCaption
941.083	Reign of George V, 1910-1936	T1--09041	1900-1919
941.083	Reign of George V, 1910-1936	T1--09042	1920-1929
941.083	Reign of George V, 1910-1936	T1--09043	1930-1939

Table 7.3: Assignment of Multiple Time Facet Codes to Schedule Caption

Table 7.3 illustrates how, if we work just with the Dewey categories, three additional Time facet codes could be assigned to the historical period span 1910-1936. However, following examination of library item titles assigned to codes such as this, crossing decades or centuries, it was considered that it would be confusing for the user and inaccurate if, for example, a library item relating just to the first part of the reign of George V was also included when a user selected the caption 1930-1939, as the following sample titles demonstrate:

LibraryItemID	Title
110815	British politics in the 1930s and 1940s
308708	Baldwin and the Conservative Party : the crisis of 1929-1931
27484	Men and power, 1917-1918
162344	The decline and Fall of Lloyd George : and great was the fall thereof
120468	Wars and welfare : Britain 1914-1945
200062	The peers, the parties and the people : the general elections of 1910
76035	Britain in the nineteen thirties
76269	Britain in the Nineteen Twenties
31293	Amateurs and professionals in British politics, 1918-59

Table 7.4: Sample Library Item Titles Relating to the Schedule Caption 1910-1936

As Table 7.4 shows, the above titles relate to a particular part of the historical period span 1910-1936 and beyond, in some cases.

As a result, in addition to the classification codes, the library items were also examined to ascertain the most appropriate Time facet code. The following general principles were used to apply a single 'best fit' TimeFacetID code:

- If an historical period code spans three or more time facet (Standard Subdivision) codes then assign the more general code. The above 1910-1936 example would be assigned the code *T1--0904 20th century, 1900-1999*.
- If just two time facet codes are applicable then use the earliest of these or the latter if it is the most predominant time period and is justified by the nature of the library item titles.

Despite this, a few false drops will undoubtedly occur, some of which have been caused by the item's original classification. For example, the code *941.084 1936-1945* had seemingly misclassified library item titles such as:

LibraryItemID	Title
226663	Never again : Britain 1945-1951
64917	Labour in power 1945-1951
199670	The Labour governments, 1945-51
188009	The politics of British defense policy, 1945-1962

Table 7.5: Possible Misclassified Items within Dewey Code 941.084, 1936-1945

Although only judging by the titles, the above examples would have perhaps been more suitably classified under *941.085 1945-1999*. Another example discovered to cause scatter and potential confusion for the user is when broad and close classification has been used simultaneously for the same subject area, the lack of consistency having the adverse effect of unnecessarily dividing rather than collating the library collection. For example, the 210 library items classed under *941.08 Period of Victoria and House of Windsor, 1837-* (potentially able to 'contain' library items from 1837 to the present day) overwhelmingly included only library items relating to the period of Victoria yet with another 197 contained in the more specific (and more appropriate) *941.08 Reign of Victoria, 1837-1901*. Perhaps this separation is an

historical legacy whereby in earlier DDC editions 941.08 once referred exclusively to the Victorian Period.

In summary, the 169 library item Dewey Codes, each having more than 10 library items, were matched against the library collection on the Assignable (Enumerated Schedule) Dewey Code field which retrieved 2670 library items, all with a time facet.

7.2.4.2 Standard Subdivisions

Another source of time-related codes is those synthesized from the Standard Subdivisions (Table 1). As discussed in section 7.2.2, Standard Subdivisions can be added either with or without instruction to any number where the subject in question approximates the whole of the number and unless there are instructions otherwise. The Standard Subdivision time notation comprises the codes 0901-0905, the 09-*Historical, geographic, persons treatment* prefix being a good facet indicator for unambiguous identification.

To identify library items with Standard Subdivision time codes, a query was constructed with the criteria:

*Like "???.*090[1-5]*"*

As can be seen from the above criteria formulation, the wildcard * character preceding the 09 and after the square brackets also identifies irregular Standard Subdivision notations with multiple zeros or 0901-0905 codes preceded or followed by other notations. This query retrieved 548 library items, comprising 359 distinct built Dewey codes.

LibraryItemID	DeweyCode	Title	Table1TimeCode
44	746.9209034	The fashionable lady in the 19th century	T1--09034
957	746.9209042	Women's dress in the Jazz Age	T1--09042
6721	320.540904	Nationalism in the twentieth century	T1--0904
53693	746.0902	Textiles and clothing : c. 1150-c. 1450	T1--0902
64884	920.00904	A Dictionary of twentieth century world biography	T1--0904
114769	133.50901	Origins of astrology	T1--0901

Table 7.6: Example of Library Items with Standard Subdivision Time Codes 0901-0905

As exemplified in Table 7.6, the Time Standard Subdivision 0901-0905 codes invariably comprises the last 4 or 5 digits of the Dewey Code; it may directly follow the decimal point or be prefixed by other digits. Embedded within the library item code in its current form, will not allow direct mapping onto our Time Facet, therefore the 'Table1TimeCode' column, shown above, was added to the query and an expression created to re-fashion the code into the correct format, as follows:

Table1TimeCode: "T1--" & Mid$([DeweyCode],InStr([DeweyCode],"090"))

This expression serves to create an extra column named 'Table1TimeCode' with the time code extracted in same format as the Standard Subdivision time periods (and our Time Period Facet), firstly by identifying the position of the text string "090" onwards in the DeweyCode field (the InStr expression) and subsequently extracting this using a Mid$ expression, prefaced by "T1--".

7.2.4.3 Main Class and Division Standard Subdivisions

Main class codes (the top level of the classification) and divisions (the second level) both end in zero therefore when a Standard Subdivision code is added these terminal zeros are dropped. This means that the decimal point is in a different place (i.e. amid the 0901-5 notation) compared to those extracted and discussed in the previous section, where the 0901-5 notations appear *after* the decimal point. Consequently, the pattern-matching query described in the previous section would not retrieve Main Class and Division Standard Subdivisions, hence their extraction by separate queries described below.

The four Main Classes that relate to historical treatment were incorporated into the query, namely 109, 509, 609 and 709. The excluded notations 009, 209, 309 are unassigned, *409 Geographic and persons treatment* does not include historical treatment, 809 is covered in a separate retrieval of literature time codes (see section 7.2.4.5) and 909 is assigned to the subject of world history.

A query was created to identify library items having such Main Class and Standard Subdivision time codes, with the following criteria:

Like "109.0[1-5]" Or Like "[5-7]09.0[1-5]*"*

This query identified 455 library items.

LibraryItemID	Title	DeweyCode	AssignableDeweyCode	Table1TimeCode
122690	The nineteenth century	509.034	509	T1--09034
128458	The scientific revolution	509.032	509	T1--09032
129231	20th century discovery	509.04	509	T1--0904
153114	Six wings : men of science in the Renaissance	509.031	509	T1--09031
173661	Topics in the history of mathematics	509.032	509	T1--09032
105326	Art of the Celts	709.01	709.01	T1--0901
105886	The art of ancient Peru	709.010985	709.01	T1--0901 0985
267571	Surrealism and after : the Gabrielle Keiller collection	709.04063	709.04063	T1--0904063
272498	Performance art : into the 90s	709.04074	709.04074	T1--0904074

Table 7.7: Main Class Standard Subdivision Library Items and Extracted Time Codes

As described previously, following identification of the library item time codes, the 0901-5 element needs extracting and prefacing by 'T1--' so as to match the codes in our Time Facet. A mid string expression was added to the query to achieve this, as follows:

Table1TimeCode: "T1--09" & Mid$([DeweyCode],5)

This expression extracts the notation from the fifth character onwards and precedes this by "T1--09". Table 7.7 shows columns for both the Dewey code and the Assignable Dewey code. The Dewey code contains both the Schedule and built elements, where occurring, of the code; whereas the Assignable Dewey code contains just the part which appears in the enumerated Schedules, hence if a code has been extended, the extension will be removed to leave just the Schedule component.

The above expression extracts from the Dewey Code field and, as can be seen in Table 7.7, this has caused problems for some notations, namely the 709s. Unlike the other 09 Main Classes included in the query, the 709 historical notations are enumerated in the Schedules, although would not have been extracted with the

Schedule notations (see Section 7.2.4.1) as they have captions such as *Surrealism* where the time-related concept is not explicit, hence their treatment here. For example, in Table 7.7, the item entitled *The art of ancient Peru* has been extended beyond the Schedule component (*709.01 Arts of nonliterate peoples, and earliest times to 499*) with an area notation 0985 for Peru. The time code, if extracted from the Dewey code version of this library item code as shown in the Table1TimeCode field, would not match the codes in our time facet; however, it would if extracted from the **709.01** Assignable version, since the time element is already incorporated. Conversely, as shown in Table 7.7, the 509 codes *need* their built element (in fact, their time component) included in order to match and thus this query needs to be based on the Dewey code.

Due to this discrepancy, the 709 criteria was removed from the previous query and subsequently incorporated into another query based on the Assignable Dewey code. This reduced the previous result set based on the Dewey code field from 455 to just 23 library items.

A new query was created for the 709s with the criteria:

Like "709.0[1-5]"*

Together with the mid string expression, retrieving 432 library item codes:

Table1TimeCode: "T1--09" & Mid$([AssignableDeweyCode],5)

However, examination of the retrieved codes showed that there were extra codes not actually present in the Time Facet but reduction by 1 or 2 digits would achieve a match. For example, an item entitled 'Neoclassicism' with the Dewey code *709.0341 Classical revival (Neoclassicism)*, which does not appear in the Time Facet, could be reduced to the broader code *709.034 19th century, 1800-1899.*

Before applying such broader codes, firstly it was established which codes did indeed match exactly; to do this, the Table1TimeCode in the previous query was joined to the TimeFacetID in the Time Facet table, identifying 360 library items. The remaining

non-matching codes were then incorporated into two further queries both based on the Table1TimeCode field of the previous 709 query, the first with the criteria:

*Like "*09011*" Or Like "*09040*"*

And the second with:

*Like "*09021?" Or Like "*09033?" Or Like "*09034?"*

The need for two queries was due to the fact that some codes needed the first 8 characters extracting in order to match the Time Facet codes whilst the others needed the first 9. The appropriate string expression was incorporated into each query:

Table1TimeCode2: Left$([Table1TimeCode],8)
Table1TimeCode2: Left$([Table1TimeCode],9)

These queries retrieved 65 and just 7 library items respectively.

The Divisions (second level hierarchy) with Standard Subdivisions were easier to extract than the Main classes. A query was constructed with the following criteria:

Like "[0-7]?0.90[1-5]"*

The 800 literature class is not included as it is dealt with separately in section 7.2.4.5. The time element was extracted using the expression:

Table1TimeCode: "T1--0" & Mid$([DeweyCode],5,4)

This mid-string expression, starting from the fifth character, extracts this and the next three characters, preceded by "T1--0", for example:

LibraryItemID	Title	DeweyCode	Table1TimeCode
187919	Magazines in the twentieth century	050.904	T1--0904
112949	Seventeenth-century metaphysics : an examination of some main concepts and theories	110.9032	T1--09032
83398	Psychology : the science of mental life	150.904	T1--0904
143920	Psychology : the science of mental life	150.904	T1--0904

Table 7.8: Division Standard Subdivision Library Items and Extracted Time Codes

This query identified 287 library items. The Table1TimeCodes from this query were then matched against the Time Facet codes to eliminate any false drops; there was only one:

LibraryItemID	Title	DeweyCode	Table1TimeCode
117070	Economic history and the social sciences : problems of methodology	330.9018	T1--09018

Table 7.9: 'False Drops' Retrieved by the Division Standard Subdivision Time Query

The Main Class and Divisions with Standard Subdivision time codes therefore comprised 741 library items in total (23 + 360 + 65 + 7 + 286), which combined with the previous 548 regular Standard Subdivisions made a total of 1289 library items with Standard Subdivision time facet notations.

7.2.4.4 Combining sub-sets

Union query 1: Combining Schedule and Standard Subdivision Library Items

As described in the previous sections, these Schedule and Standard Subdivision library items are currently retrieved using a series of s*eparate* queries; a single union select query was therefore created using these to combine the sets and eliminate duplicates.

Seven queries were combined:

Time Facet Query	Number of Library Items
Schedules	7670
Regular Standard Subdivisions	548
Main Class Standard Subdivisions (without 709s)	23
709 Main Class Standard Subdivisions matching Time Facet	360
709 Unmatching 1	65
709 Unmatching 2	7
Division Standard Subdivisions	286
TOTAL	8959

Table 7.10: Combination of Standard Subdivision Time Queries

There were no duplicates amongst the retrieved library items from the individual queries as the sum total for these was found to be identical to that of the union query.

7.2.4.5 Literary Time Periods

Time period is an important and ubiquitous component in the Dewey 800 Literature Main class, making it a fertile source for populating the time facet. Unlike other Main Classes, it displays very little enumeration, with the majority of codes being built following a series of, rather complicated, instructions using three auxiliary tables:

3A - for works by or about individual authors
3B - for works by or about more than one author
3C - for works by or about more than one author, added under instruction from Table 3B and notations 808-809

Tables 3A and 3B essentially contain literary form notations consisting of a single digit, e.g. 1 = poetry, whilst Table 3C refers to other characteristics such as theme or kind of person. The time notation is added for 801-809 literature generalia using the 01-05 Standard Subdivisions whereas separate historical period tables in the Schedules are used for each language in the 810-899 span.

Described by Chan et al (1996, p.128) as "one of the most faceted of all the main classes", the various citation orders for the literature class are well documented in Chan's guide and the DDC Manual. Therefore, unlike our previous deconstruction efforts based on pattern matching of the required code component regardless of

position after the decimal point (due to the fluidity of this anyway) and subsequent exclusion of false drops, the syntax or citation order of the Dewey Code was the primary tool for identifying library items with a time component.

The citation orders for the literature class constitute slight variations on the basic formula:

language + literary form or kind of form + period

Knowing exactly where within the Dewey code a time component would occur led to the creation of a series of queries founded upon positional criteria. The first set of queries identifies library items classified as literature generalia with a time component, and the second, much larger set, retrieves items classified under each distinct language as being by or about individual authors or anthologies/criticism, again all with a time facet component.

Once the library item Dewey codes had been identified using these queries, criteria to extract the Time element of the code needed to be incorporated into the query in order that it may be mapped onto the respective code in the time facet, i.e. T1--0901-0905. In the case of the language based literature time codes, transitional tables were created for each language, mapping the distinctive language time ID against the time facet ID. For example, the digit *3* in the French literary period table is *Renaissance period, 1500-1599*, which would be converted into the time facet (Table 1/Standard Subdivision) code *T1--09031 16th century, 1500-1599*.

The final stage entails combining the retrieved sets using union queries. The procedure is now described in more detail, beginning with literature generalia, followed by literature by language - individual authors and anthologies/criticism.

Literature Generalia

Literature generalia comprises Standard Subdivisions of literature 801-807, followed by *808.001-808.7 Rhetoric* and *808.8-809 collections of literary texts* (anthologies and criticism respectively) in three or more languages from different language families. As Standard Subdivisions, the notations 801-807 would not usually be given an additional Standard Subdivision time notation; however, 801.95 philosophy of

literary criticism contains the history of this topic, with 0901-0905 Standard Subdivision time notations included in Relative Index entries, e.g. *801.950**904*** a built number representing literary criticism in the 20th century. Since such codes have their time element appended via the Standard Subdivisions in the regular way, library items with these codes will have been retrieved by queries described in section 6.1.4.2 Standard Subdivisions so will not be discussed further here. Additionally, as not constituting a Standard Subdivision, the rhetoric span 808.001-808.7 is subject to Standard Subdivision time codes, which will also be retrieved in the same Standard Subdivision queries. Consequently, the focus of this section is the identification and extraction of the time element in the remaining library item codes *808.8 Collections of literary texts from more than two literatures* and *809 History, description, critical appraisal of more than two literatures.*

Chan et al (1996, pp.136-140) and the DDC manual (OCLC, 1998) detail the various citation orders for general literary collections. From this, three were selected as including the time period facet:

1) 808.80/809 Base number + Time period 01-05 (one facet)
2) 808.80/809 Base number + Time period 01-05 or Feature/Theme/Persons Table 3C notations (one facet)
3) 808.8/809 Base number + Form + Time period 01-05 (two facets)

As the above demonstrates, the time period notations for literary collections are taken from the Standard Subdivisions but without the 09 prefix (hence their exclusion from the Standard Subdivision queries described earlier); however, the addition of this will enable a direct mapping onto our Time Facet. It can be seen that the second citation order has two options available, either a time period facet or a topical feature/theme/persons facet, depending on emphasis in the library item. If time period is emphasized then feature/theme/persons does not apply and vice versa and as such the syntax is identical to that of the first above citation orders, 1) base number and time period. Another point of note is the difference in 808.8 base numbers; namely, the missing extra zero in 3), the form and time period facet combination.

Queries were created based on these citation orders to identify relevant library items, which will now be described in turn.

Base Number and Time Period

Separate queries were constructed to identify relevant 808.8 and 809 library item codes with time period facets, as detailed in Table 7.11:

Dewey code	Caption	Query criteria	Expression	No. library items
808.8	*Collections of literary texts from more than two literatures*	Like "808.800[1-5]*"	TimePeriodID: "T1--090" & Mid$([DeweyCode],8)	0
809	*History, description, critical appraisal of more than two literatures*	Like "809.0[1-5]*"	TimePeriodID: "T1--090" & Mid([DeweyCode],6)	81 77 (Matching Time Facet)

Table 7.11: Literature Base Number 808.8 and 809 and Time Period Library Items

Table 7.11 illustrates the query criteria to identify 808.8 and 809 library items with a time facet and the expressions employed to extract this particular component. As shown, there were no library items classified as 808.8 with a time facet, yet the 809 query initially identified 81 library items; this query was subsequently matched to the Time Facet using an inner join query, limiting the set to 77 library item codes.

Base Number and Time period 01-05 or Feature/Theme/Persons

As discussed in the previous section, this second citation order is identical to that described in the above section and thus does not necessitate an additional set of queries.

Base Number, Literary Form and Time Period

As in the above 'Base Number and Time Period' section, queries were created for both 808.8 and 809, detailed in Table 7.12.

Dewey code	Caption	Query criteria	Expression	Library items
808.8	*Collections of literary texts from more than two literatures*	Like "808.8[!0]*0[1-5]*"	TimePeriodID: "T1--09" & Mid$([DeweyCode],7)	1
809	*History, description, critical appraisal of more than two literatures*	Like "809.[!0]*0[1-5]*"	TimePeriodID: "T1--09" & Mid$([DeweyCode],6)	103 86 (Matching Time Facet)

Table 7.12: Literature Base Number 808.8 and 809, Literary Form and Time Period Library Items

Table 7.12 shows that there was 1 library item classified at 808.8 with a time facet and 86 library items at 809. The [!0]* element in the query criteria means that any character except zero will be included directly after the decimal point; a zero is not wanted since this will mean that that the library item code is lacking a literary form facet (notations 1-8 from Table 3A and 3B). As in Table 7.11, the 809 code query contained a small number of false drops (17 in total) which were removed when the initial query was matched to the time facet table by way of an inner join query, leaving 86 library item codes.

Combing Subsets

Union query 2: Combining Literature Generalia Library Items

The second union query combines the four queries described in the previous sections:

Query	Number of Library Items
808.80 + Time period	0
809 + Time period	77
808.8 + Form + time period	1
809 + form + time period	86
TOTAL	**164**

Table 7.13: Combination of Literature Generalia Time Queries

The union query retrieved the same number of library items as the sum of the individual queries so again there were no duplicate library items amongst the individual queries.

Literature by language

Literature that relates to or is written in a particular language is classified with that language using notations 810-899, with the second digit designating the specific language or language family facet. As described in the introduction to section 7.2.4.5, each individual language has a separate literary period table, which led to the manual creation of a series of transitional tables, mapping each set of language time period notations found in the Schedules onto the Standard Subdivision Time Facet notations 0901-0905, for example:

LanguageTime PeriodID	LanguageTimePeriodCaption	TimeFacetID	TimeFacetCaption
1	*Colonial period, 1607-1776*	T1--09032	*17th century, 1600-1699*
2	*1776-1829*	T1--09033	*18th century, 1700-1799*
3	*1830-1861*	T1--09034	*19th century, 1800-1899*
4	*1861-1899*	T1--09034	*19th century, 1800-1899*
5	*1900-1999*	T1--0904	*20th century, 1900-1999*
52	*1900-1945*	X1--190045	*1900-1945*
54	*1945-1999*	X1--194599	*1945-1999*
6	*2000-*	T1--0905	*21st century, 2000-2099*

Table 7.14: American literature in English, Time Period Translation Table

Table 7.14 shows an example translation table; it should be noted that in the database only the LanguageTimePeriodID and TimeFacetID fields were included and the captions are only displayed here to inform the description of the translation process. Examining the above table captions, it can be seen that frequently the language time period notations cut across centuries so can only be provided with an approximate 'best fit' Time Facet notation, for example *Colonial period 1607-1776* translates to *17th century, 1600-1699*.[10] Unfortunately, this problem occurred when translating many of the language period tables. As an aid to increasing the specificity of subject access to the University of Huddersfield's predominantly twentieth century literature collection, the Standard Subdivision codes used in our Time Facet were further subdivided into early and latter parts of the twentieth century, viz.1900-45 and 1945-99, a common division in the language literary period tables. Twenty-three

[10] For reasons discussed earlier in section 6.1.4.1, only one translation code is assigned to each language time period.

transitional tables were created for all the languages within notations 810-889, all European languages; this might seem a rather large number for 8 DDC Divisions but this is due to the fact that certain languages (viz. German, French, Spanish, Greek and Italian) have an 8?9 'other languages' Section which contains further period tables.

The only division that has not been tackled is the 890 Literature of Other Languages. This subdivision is hierarchically deep and extremely diverse containing literatures from 'East Indo-European and Celtic literatures' to 'Literatures of non-Austronesian languages of Oceania, of Austronesian languages, of miscellaneous languages' and 33 individual period tables. The huge number of seemingly unrelated literatures bunched together in this catchall class betrays the scheme's 19th century roots. As Chan et al (1996, p.129) explain:

> The literatures of individual languages are classed in 810-899. Each language that at one time was considered to be of importance to western scholars has been allotted a division in 810-880. Consequently, many other important literatures are classed as subdivisions of 890 Literatures of other languages, and thus have long numbers; still other national languages receive only scant treatment.

Moreover, period notation is not added to works for the majority of those writers who do not live in the ancestral continent of their mother tongue, for example, periods for French literature does not apply to authors in Africa or Canada; Australian authors do not have English periods assigned nor Spanish periods for Mexico. This is branded by Chan et al (1996) as "a serious shortcoming in Dewey that cannot be rectified soon, because it would require major restructuring of the 800s" (p.129).

Consequently, it was decided that owing to this complexity and the relatively few library items (314) with this 89? code, deconstruction to identify and extract time period elements, although possible, would be heavily time intensive and thus not feasible in cost-benefit terms. One way of reducing the effort would be to construct a query to deconstruct just those 89? codes in the library's current holdings but this would not allow for new additions with 89? codes other than these.

The subsequent sections describe how each set of library item literature by language Dewey codes were identified and deconstructed, firstly in terms of those by and about individual authors and secondly, anthologies and criticism of literature with a time facet.

Individual Authors

For all languages, works by and about individual authors have the following facets and citation order:

1) Main Class + language + literary form + **period**

Miscellaneous writings (8?8) of a particular language are subject to an additional facet:

2) Main Class + language + 8 + **period** + special form

The main class, language and period components are all obtained from the Dewey Schedules, whilst the literary form facet is added from Table 3A; for example the novel Emily Brontë's "Wuthering Heights" would be classed as 8 (Main Class) + 2 (English Literature) + 3 (Fiction) + 8 (Victorian period, 1837-1899) = 823.8

Thirteen queries were created for each of the 810-889 individual literatures using an identical query design but with language-specific criteria, as illustrated in the table below.

Language	Query criteria	Expression	Number of library items
American	Like "81?.[1-6]" Or Like "81?.52" Or Like "81?.54"	TimePeriodID: Mid$([DeweyCode],5)	852
English	Like "82[0-8].[0-9]" Or Like "82[0-8].91" Or Like "82[0-8].912" Or Like "82[0-8].914" Or Like "82[0-8].92"	TimePeriodID: Mid$([DeweyCode],5)	5848
French	Like "84[0-8].[1-9]" Or Like "84[0-8].91" Or Like "84[0-8].912" Or Like "84[0-8].914" Or Like "84[0-8].92"	TimePeriodID: Mid$([DeweyCode],5)	2828
German	Like "83[0-8].[1-9]" Or Like "83[0-8].21" Or Like "83[0-8].22" Or Like "83[0-8].91" Or Like "83[0-8].914" Or Like "83[0-8].92"	TimePeriodID: Mid$([DeweyCode],5)	1245
Italian	Like "85[0-8].[1-9]" Or Like "85[0-8].91" Or Like "85[0-8].912" Or Like "85[0-8].914" Or Like "85[0-8].92"	TimePeriodID: Mid$([DeweyCode],5)	41
Spanish	Like "86[0-8].[1-7]" Or Like "86[0-8].62" Or Like "86[0-8].64"	TimePeriodID: Mid$([DeweyCode],5)	86
Latin	Like "87[0-8].[1-4]"	TimePeriodID: Mid$([DeweyCode],5)	6
Greek	Like "88[0-8].[1-3]"	TimePeriodID: Mid$([DeweyCode],5)	3
Other French - Catalan Provencal	Like "849.[0-8][1-6]" Or Like "849.[0-8]52" Or Like "849.[0-8]54"	TimePeriodID: Mid$([DeweyCode],6)	0
Other Germanic -	Like "839.1[0-8][1-4]" Or Like	TimePeriodID:	75

Yiddish Frisian Dutch Afrikaans Low German Modern Icelandic Swedish Danish Norwegian New Norwegian	"839.2[0-8][1-5]" Or Like "839.31[0-8][1-7] " Or Like "839.31[0-8]62" Or Like "839.31[0-8]64" Or Like "839.36[0-8][1-6]" Or Like "839.4[0-8][1-4]" Or Like "839.69[0-8][1-5]" Or Like "839.7[0-8][1-8]" Or Like "839.8[1-3][0-8][1-8]" Or Like "839.8[1-3][0-8]72" Or Like "839.8[1-3][0-8]74"	IIf(Left$([DeweyCode],5) <>"839.8",Mid$([Dewey Code],7),Mid$([DeweyCo de],8))	
Other Greek - Modern Greek	Like "889.[0-8][1-4]" Or Like "889.[0-8]32" Or Like "889.[0- 8]34"	TimePeriodID: Mid$([DeweyCode],6)	2
Other Italian - Romanian	Like "859.[0-8][1-3*]"	TimePeriodID: Mid$([DeweyCode],6)	0
Other Spanish - Portuguese	Like "869.[0-8][1-5]" Or Like "869.[0-8]41" Or Like "869.[0- 8]42"	TimePeriodID: Mid$([DeweyCode],6)	0
TOTAL			10,960

Table 7.15: Query Criteria for Individual Literatures by Language and Time Facet

Table 7.15 shows a total of 10,960 library items (4.3% of the complete library collection) identified as literature by or about individual authors with a time facet. The 'other' language codes (8?9) and captions frequently subsume more than one language and therefore period table, resulting in a more complex query criteria and expression, notably Other Germanic literatures. Not all languages have an 'other languages' section (i.e. American literature) or have period tables enumerated (Italic literatures). English is idiosyncratic in that it utilises its 829 code for Old English (Anglo Saxon) for which the time period is inherent; a separate query design was therefore used to identify such library items.

Time period does not need to be identified via citation order for Old English, thus it is not necessary to retrieve separately library items relating to individual authors or anthologies and criticism (see Table 7.15); only one query is required. A query was constructed with the criteria *Like "829*"* and the expression *TimePeriodID: "T1--09021"*, (6th-12th centuries, 500-1199), the most appropriate time period for Old English. Although not necessarily relating to individual authors, this query retrieved an additional 47 library items.

Miscellaneous individual author items, as detailed in citation order 2, have an additional special form facet. As a result of this, the citation order is almost identical to citation order 2 in Table 7.17 below and, consequently, it proved more economical

to extract such items as part of this anthology/criticism query than create a series of separate miscellaneous individual author queries. The identification of these items is therefore described in the later section 'Citation order 2: Anthologies/criticism in a particular language and literary form with a time period facet'.

Anthologies and criticism

Anthologies and criticism of literature by language with a time facet comprise four different citation orders:

1) Main Class 8 + language 1-9 + anthologies 080 or criticism 090 + **period** tables 01-09

2) Main Class 8 + language 1-9 + form 2-7 + **period** tables 01-09 + anthologies 08 or criticism 09

3) Main Class 8 + language 1-9 + anthologies 08 or criticism 09 + feature/theme/persons Table 3C + **period** Standard Subdivisions 0901-0905

4) Main Class 8 + language 1-9 + form 1-7 + **period** tables + anthologies 08 or criticism 09 + feature/theme/persons Table 3C

The above citation orders show the gradation of facets (order 1 having the least number of facets and 4 the highest) and their combinations. Note that the period notation is synthesised using the Schedule period tables for each language in all cases except citation order 3, for which the Standard Subdivision (Table 1) notations are employed and as such will match our Time Facet exactly. According to Chan et al (1996, pp.145-146) this is because in Table 3B --08 and --09 the period tables are added with two extra zeros, i.e. notations 08001-009 and 09001-009 whilst the feature/theme/persons Table 3C notations are added directly to the 08/09 codes; since there is no provision to combine period and feature/theme/persons in either Table 3B or 3C, the rule of fewest zeros applies and the feature/theme/persons notations take precedence over time period. Time period can, however, still be included but it is instead appended as a regular Standard Subdivision notation.

The following sub-sections detail how each citation order was used to identify library items and extract the time period component.

Citation order 1: Anthologies/criticism in a particular language with a time period facet

Queries were created for each of the thirteen languages, as detailed in the table below:

Language	Query criteria	Expression	Number of library items
American	Like "810.[8-9]00[1-6]" Or Like "810.[8-9]0052" Or Like "810.[8-9]0054"	TimePeriodID: Mid$([DeweyCode],8)	21
English	Like "820.[8-9]00[1-9]" Or Like "820.[8-9]0091" Or Like "820.[8-9]00912" Or Like "820.[8-9]00914" Or Like "820.[8-9]0092"	TimePeriodID: Mid$([DeweyCode],8)	218
French	Like "840.[8-9]00[1-9]" Or Like "840.[8-9]0091" Or Like "840.[8-9]00912" Or Like "840.[8-9]00914" Or Like "840.[8-9]0092"	TimePeriodID: Mid$([DeweyCode],8)	138
German	Like "830.[8-9]00[1-9]" Or Like "830.[8-9]0021" Or Like "830.[8-9]0022" Or Like "830.[8-9]0091" Or Like "830.[8-9]00914" Or Like "830.[8-9]0092"	TimePeriodID: Mid$([DeweyCode],8)	75
Italian	Like "850.[8-9]00[1-9]" Or Like "850.[8-9]0091" Or Like "850.[8-9]00912" Or Like "850.[8-9]00914" Or Like "850.[8-9]0092"	TimePeriodID: Mid$([DeweyCode],8)	1
Spanish	Like "860.[8-9]00[1-7]" Or Like "860.[8-9]0062" Or Like "860.[8-9]0064"	TimePeriodID: Mid$([DeweyCode],8)	1
Latin	Like "870.[8-9]00[1-4]"	TimePeriodID: Mid$([DeweyCode],8)	0
Greek	Like "880.[8-9]00[1-3]"	TimePeriodID: Mid$([DeweyCode],8)	0
Other French - Catalan Provencal	Like "849.0[8-9]00[1-6]" Or Like "849.0[8-9]0052" Or Like "849.0[8-9]0054"	TimePeriodID: Mid$([DeweyCode],9)	0
Other Germanic - Yiddish Frisian Dutch Afrikaans Low German Modern Icelandic Swedish Danish Norwegian New Norwegian	Like "839.10[8-9]00[1-4]" Or Like "839.20[8-9]00[1-5]" Or Like "839.310[8-9]00[1-7]" Or Like "839.310[8-9]0062" Or Like "839.310[8-9]0064" Or Like "839.360[8-9]00[1-6]" Or Like "839.40[8-9]00[1-4]" Or Like "839.690[8-9]00[1-5]" Or Like "839.70[8-9]00[1-8]" Or Like "839.70[8-9]0072" Or Like "839.70[8-9]0074" Or Like "839.8[1-3]0[8-9]00 [1-8]" Or Like "839.8[1-3]0[8-9]0072" Or Like "839.8[1-3]0[8-9]0074"	TimePeriodID: Mid$([DeweyCode],10)	0
Other Greek - Modern Greek	Like "889.0[8-9]00[1-4]" Or Like "889.0[8-9]0032" Or Like "889.0[8-9]0034"	TimePeriodID: Mid$([DeweyCode],9)	0
Other Italian - Romanian	Like "859.0[8-9]00[1-3*]"	TimePeriodID: Mid$([DeweyCode],9)	0
Other Spanish - Portuguese	Like "869.0[8-9]00[1-5]" Or Like "869.0[8-9]0041" Or Like "869.0[8-9]0042"	TimePeriodID: Mid$([DeweyCode],9)	0
TOTAL			454

Table 7.16: Query Criteria for Anthologies/Criticism Citation Order 1

Table 7.16 shows a total of 454 library items retrieved for citation order 1. For each language, the query criteria identify library items with this particular citation order and the expression then extracts just the time period component. For this citation order, the time facet is the final component of the Dewey code; for example, French literature in the 20th century has the criteria "840.[8-9]00**91**":

*8 + 4 + 0. [8-9] +0 + 0 + **91** = Literature + French + anthologies/criticism + facet indicator + facet indicator + **20th century***

The use of square brackets allows numbers within a range to be retrieved, thus the [8-9] will include 08 anthologies and 09 criticism. Similarly, in the query criteria above, ranges have been used to simplify time period identification; for example, the expression Like "820.[8-9]00[**1-9**]" will retrieve all the period table notations for English literature between 1 and 9. The majority of period tables, including English, also have period notations of more than one digit. Although simply appending an asterisk after the range to truncate it (i.e. [1-9]*) could have retrieved these, it was decided to retrieve such period notations using separate criteria (as shown in the previous 20th century English Literature example) to guard against false drops.

The query criteria and expressions in Table 7.16 indicate that for the eight main language divisions, America to Greek, the time facet begins and was extracted from the eighth character onwards, whilst for the 'other' languages it appears between one and three characters later due to the extra length of such language digits. 'Other Germanic languages', for instance, have their time facet beginning at either the tenth (e.g. 739.1 Yiddish) or eleventh digit (e.g. 739.31 Dutch). The expression shown in Table 7.16 extracts from the tenth to ensure that the entire language component is isolated, meaning that language codes represented by more than one digit, like Dutch, would have their time facet notation prefixed by zero; however in the event, this could easily be removed by a subsequent query. Another complication is that there are ten period tables contained within 'Other Germanic literatures' yet for our purposes the individual literatures have been retrieved en masse; if a large and diverse result set were retrieved then the construction of further queries to match each particular language to its correct period table would be necessary. Alternatively, if a library's holdings included a specialist collection of Germanic literature then these individual languages might merit separate treatment (i.e. separate queries) from the outset.

However, since there were no 'Other Germanic literature' library items for this citation order in the University of Huddersfield library collection, this was not pursued further.

Citation order 2: Anthologies/criticism in a particular language and literary form with a time period facet

Language	Query criteria	Expression	Number of library items
American	Like "81[1-8].[1-6]0[2-3 7-9]" Or Like "81[1-8].50[2-3 7-9]" Or Like "81[1-8].540[2-3 7-9]"	LanguageTimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),5)	74
English	Like "82[1-8].[1-9]0[2-3 7-9]" Or Like "82[1-8].910[2-3 7-9]" Or Like "82[1-8].9120[2-3 7-9]" Or Like "82[1-8].9140[2-3 7-9]" Or Like "82[1-8].920[2-3 7-9]"	LanguageTimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),5)	789
French	Like "84[1-8].[1-9]0[2-3 7-9]" Or Like "84[1-8].910[2-3 7-9]" Or Like "84[1-8].9120[2-3 7-9]" Or Like "84[1-8].9140[2-3 7-9]" Or Like "84[1-8].920[2-3 7-9]"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),5)	141
German	Like "83[1-8].[1-9]0[2-3 7-9]" Or Like "83[1-8].210[2-3 7-9]" Or Like "83[1-8].220[2-3 7-9]" Or Like "83[1-8].910[2-3 7-9]" Or Like "83[1-8].9140[2-3 7-9]" Or Like "83[1-8].920[2-3 7-9]"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),5)	75
Italian	Like "85[1-8].[1-9]0[2-3 7-9]" Or Like "85[1-8].910[2-3 7-9]" Or Like "85[1-8].9120[2-3 7-9]" Or Like "85[1-8].9140[2-3 7-9]" Or Like "85[1-8].910[2-3 7-9]"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),5)	4
Spanish	Like "86[1-8].[1-7]0[2-3 7-9]" Or Like "86[1-8].620[2-3 7-9]" Or Like "86[1-8].640[2-3 7-9]"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),5)	16
Latin	Like "87[1-8].[1-4]0[2-3 7-9]"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),5)	2
Greek	Like "88[1-8].[1-3]0[2-3 7-9]"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),5)	0
Other French - Catalan Provencal	Like "849.[1-8][1-6]0[2-3 7-9]" Or Like "849.[1-8]520[2-3 7-9]" Or Like "849.[1-8]540[2-3 7-9]"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),6)	1
Other Germanic - Yiddish	Like "839.1[1-8][1-4]0[2-3 7-9]" Or Like "839.2[1-	TimePeriodID: Mid$(Left$([DeweyCode],Len([D	0

Frisian Dutch Afrikaans Low German Modern Icelandic Swedish Danish Norwegian New Norwegian	8][1-5]0[2-3 7-9]" Or Like "839.31[1-8][1-7]0[2-3 7-9]" Or Like "839.31[1-8]620[2-3 7-9]" Or Like "839.31[1-8]640[2-3 7-9]" Or Like "839.36[1-8][1-6]0[2-3 7-9]" Or Like "839.4[1-8][1-4]0[2-3 7-9]" Or Like "839.69[1-8][1-5]0[2-3 7-9]" Or Like "839.7[1-8][1-8]0[2-3 7-9]" Or Like "839.7[1-8]720[2-3 7-9]" Or Like "839.7[1-8][1-8]740[2-3 7-9]" Or Like "839.8[1-3][1-8][1-8]0[2-3 7-9]" Or Like "839.8[1-3][1-8]720[2-3 7-9]" Or Like "839.8[1-3][1-8]740[2-3 7-9]"	eweyCode])-2),7)	
Other Greek - Modern Greek	Like "889.[1-8][1-4]0[2-3 7-9]" Or Like "889.[1-8]320[2-3 7-9]" Or Like "889.[1-8]340[2-3 7-9]"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),6)	0
Other Italian - Romanian	Like "859.[1-8][1-3*]0[2-3 7-9]"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),6)	0
Other Spanish - Portuguese	Like "869.[1-8][1-5*]0[2-3 7-9]" Or Like "869.[1-8]410[2-3 7-9]" Or Like "869.[1-8]420[2-3 7-9]"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),5)	0
TOTAL			1102

Table 7.17: Query Criteria for Anthologies/Criticism Citation Order 2

Table 7.17 shows a total of 1102 library items with either citation order 2 or miscellaneous individual authors (as mentioned in the previous 'Individual Authors' section). In both citation orders, the time period component is embedded within the library item code, preceded by the language and literary form facets and followed by further subdivisions. Thus, if we use the same example as previously, to retrieve French literature in the 20th century in the case of both citation orders the query criteria comprises *"84[1-8].910[2-3 7-9]"*. The query criteria will retrieve library items representing both citation orders, as follows:

Citation order 2: Anthologies/criticism in a particular language and literary form with a time period facet

8 + 4 + [1-8] + ***91*** *+ 08 or 09 = Literature + French + literary form +* ***20th century*** *+ anthologies/criticism*

Individual authors of miscellaneous writings in a particular language

*8 + 4 + [1-8] + **91** + 02 or 03 or 07 or 08 or 09 = Literature + French + **20th century** + Special form*

Therefore, the difference between the above two citation orders is the particular subdivisions appended to the time period facet, all of which are encompassed by the range criteria *[2-3 7-9]*.

As shown by the expressions in Table 7.17, the time period component typically begins at the fifth character; however a simple Mid$ query, as used previously, would not only extract the time period facet but also the following subdivisions. Consequently, further criteria, *Left$([DeweyCode],Len([DeweyCode])-2*, were required to remove these final two digits, leaving a compatible time period notation.

Citation order 3: Anthologies/criticism in a particular language relating to a specific feature/theme/persons and a time period

This citation order retrieved a disappointingly small result set, with just 15 library items in total. As detailed in the introduction to the 'Anthologies/Criticism' section, this citation order is unusual in that it takes its period notation from the Standard Subdivisions rather than the language-specific period tables in the Schedules.

Since no period tables are utilised, the query criteria is more consistent across languages, entailing just a change in language digit(s) for each query, for example for English literature: *Like "820.[8-9]***090***"*

*8 Literature + 2 English + 0[8-9] Anthology or Criticism + * Feature/Theme/Person notation + **090* Period***

As can be seen from the above, the period notation appears as the final component in the Dewey code. The expression created to extract this component is based upon identification of the 090 In String and its subsequent extraction using a Mid String expression:

Table1TimeCode: "T1--" & Mid$([DeweyCode],InStr([DeweyCode],"090"))

Citation order 4: Anthologies/criticism in a particular language and literary form relating to a specific feature/theme/persons and a time period

Language	Query criteria	Expression	Number of library items
American	Like "81[1-8].[1-6]0[8-9]*" Or Like "81[1-8].50[8-9]*" Or Like "81[1-8].540[8-9]*"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),5,3)	88
English	Like "82[1-8].[1-9]0[8-9]*" Or Like "82[1-8].910[8-9]*" Or Like "82[1-8].9120[8-9]*" Or Like "82[1-8].9140[8-9]*" Or Like "82[1-8].920[8-9]*"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),5,3)	895
French	Like "84[1-8].[1-9]0[8-9]*" Or Like "84[1-8].910[8-9]*" Or Like "84[1-8].9120[8-9]*" Or Like "84[1-8].9140[8-9]*" Or Like "84[1-8].920[8-9]*"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),5,3)	146
German	Like "83[1-8].[1-9]0[8-9]*" Or Like "83[1-8].210[8-9]*" Or Like "83[1-8].220[8-9]*" Or Like "83[1-8].910[8-9]*" Or Like "83[1-8].9140[8-9]*" Or Like "83[1-8].920[8-9]*"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),5,3)	75
Spanish	Like "86[1-8].[1-7]0[8-9]*" Or Like "86[1-8].620[8-9]*" Or Like "86[1-8].640[8-9]*"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),5,3)	16
Latin	Like "87[1-8].[1-4]0[8-9]*"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),5,3)	2
Greek	Like "88[1-8].[1-3]0[8-9]*"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),5,3)	0
Other French - Catalan Provencal	Like "849.[1-8][1-6]0[8-9]*" Or Like "849.[1-8]520[8-9]*" Or Like "849.[1-8]540[8-9]*"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),6,3)	1
Other Germanic - Yiddish Frisian Dutch Afrikaans Low German Modern Icelandic Swedish Danish Norwegian New Norwegian	Like "839.1[1-8][1-4]0[8-9]*" Or Like "839.2[1-8][1-5]0[8-9]*" Or Like "839.31[1-8][1-7]0[8-9]*" Or Like "839.31[1-8]620[8-9]*" Or Like "839.31[1-8]640[8-9]*" Or Like "839.36[1-8][1-6]0[8-9]*" Or Like "839.4[1-8][1-4]0[8-9]*" Or Like "839.69[1-8][1-5]0[8-9]*" Or Like "839.7[1-8][1-8]0[8-9]*" Or Like "839.72[1-8][1-8]0[8-9]*" Or Like "839.74[1-8][1-8]0[8-9]*" Or Like "839.8[1-3][1-8][1-8]0[8-9]" Or Like "839.8[1-3][1-8]720[8-9]*" Or Like	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),7,3)	0

	"839.8[1-3][1-8]740[8-9]*"		
Other Greek - Modern Greek	Like "889.[1-8][1-4]0[8-9]*" Or Like "889.[1-8]320[8-9]*" Or Like "889.[1-8]340[8-9]*"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),6,3)	0
Other Italian - Romanian	Like "859.[1-8][1-3*]0[8-9]*"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),6,3)	0
Other Spanish - Portuguese	Like "869.[1-8][1-5*]0[8-9]*" Or Like "869.[1-8]410[8-9]*" Or Like "869.[1-8]420[8-9]*"	TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),6,3)	0
TOTAL			1227

Table 7.18: Query Criteria for Anthologies/Criticism Citation Order 4

Table 7.18 shows a total of 1227 library items following the citation order Main class + language + literary form + **period** + anthologies/criticism + feature/theme/persons. This citation order is very similar to that of citation order 2 with language, literary form and time period facets, the only difference being the ultimate period/feature/theme component. As such, the query criteria utilised to identify citation order 4 library items may also retrieve those following citation order 2, in other words library items may not necessarily have the period/feature/theme facet after the 08/09 notation. This was generally found to be the case when the queries were run, with most notations concluding with 08 or 09.

The time period notation directly follows the decimal point and an expression was created to extract this, following the format:

TimePeriodID: Mid$(Left$([DeweyCode],Len([DeweyCode])-2),5,3)

This expression creates a new TimePeriodID column, removes the final 2 characters (typically 08/09) from the Dewey code using a Left$ and then extracts a Mid$ of 3 characters (the maximum time code length) from the fifth character onwards. The result sets comprised time codes from library item notations without a period/feature/theme notation, which will exactly match the particular time period translation table, and those from library items with a period/feature/theme notation, which requires the removal of additional characters to enable a match. Consequently, four further queries were created for each language.

The first query is simply an Inner Join between the original citation order 4 query and the appropriate language translation table to find exact match time codes. A second query performed a Left Join of the original query again against the relevant language translation but with the criteria *LanguageTimePeriodID* (i.e. language time period notation) *"Is Null"*. Consequently, this selected library item codes that did not match the time period codes in the language translation tables. These 'unmatching' result sets displayed three forms of time codes with extraneous digits: ?0?, ??0 and ?0. Thus, using the 'un-matching' query, a third query was created which specified *?0?* as the query criteria in the TimePeriodID field, together with the expression *TimePeriodID2: Left$([TimePeriodID],1)* create a second TimePeriodID field and extract just the first character (i.e. the matching language time period notation) from the original TimePeriodID notation. The two remaining 'un-matching' formats ??0 and ?0 were resolved using one query, since they both end in zero. As in the third, the fourth query is based on the 'un-matching' query but with the criteria *Like "??0" Or Like "?0"* and the expression:

TimePeriodID2:IIf(Right$([TimePeriodID],1)="0",Left$([TimePeriodID],Len([TimePeriodID])-1))

This expression again creates a second TimePeriodID field and, if the original TimePeriodID ends in zero, it extracts the TimePeriod ID using a Left$ and removes the extraneous zero to leave just a 'matching' time period notation.

To exemplify the above four queries, the original English literature citation order 4 query retrieved 895 library items, 791 (88%) of which exactly matched the English time period translation table with 104 un-matching. Of these 104 un-matching time codes, 76 had the format *?0?* and the remaining 28 the format *??0* or *?0*.

Combining Subsets

Union query 3: Combining the citations orders and mapping period table codes to the Time Facet

With the exception of citation order 3, the result sets for the literature by language queries described above have all comprised library items with a language specific

time period notation. These time period codes require translation to the appropriate Time Facet (Table 1/Standard Subdivision) notation. For each language, the following three-step process was used to achieve this:

1. A union query is created to combine the queries of citation orders 1, 2 and 4, for example:

```
SELECT * FROM qryTimeLiteratureEnglish08900Collections1Facet
UNION
SELECT * FROM
qryTimeLiteratureEnglishAnthologiesCriticismFormAndPeriodFacets
UNION
SELECT * FROM qryTimeLiteratureEnglishAnthologiesCritWithAllFacetsMatching
UNION
SELECT * FROM
qryTimeLiteratureEnglishAnthologiesCritWithAllFacetsUnmatching1
UNION
SELECT * FROM
qryTimeLiteratureEnglishAnthologiesCritWithAllFacetsUnmatching2
UNION
SELECT * FROM qryTimeLiteratureEnglishIndividualAuthors;
```

With a sample result set:

LibraryItemID	TimePeriodID	Origin
141	91	Built
380	7	Built
396	91	Built
397	91	Built

Table 7.19: Sample Results for Literature by Language Union Query

2. The language-specific TimePeriodID field from the above union query is then joined with the equivalent language TimePeriodID using the appropriate language translation table but with the TimeFacetID (Table1/Standard Subdivision) field selected for display:

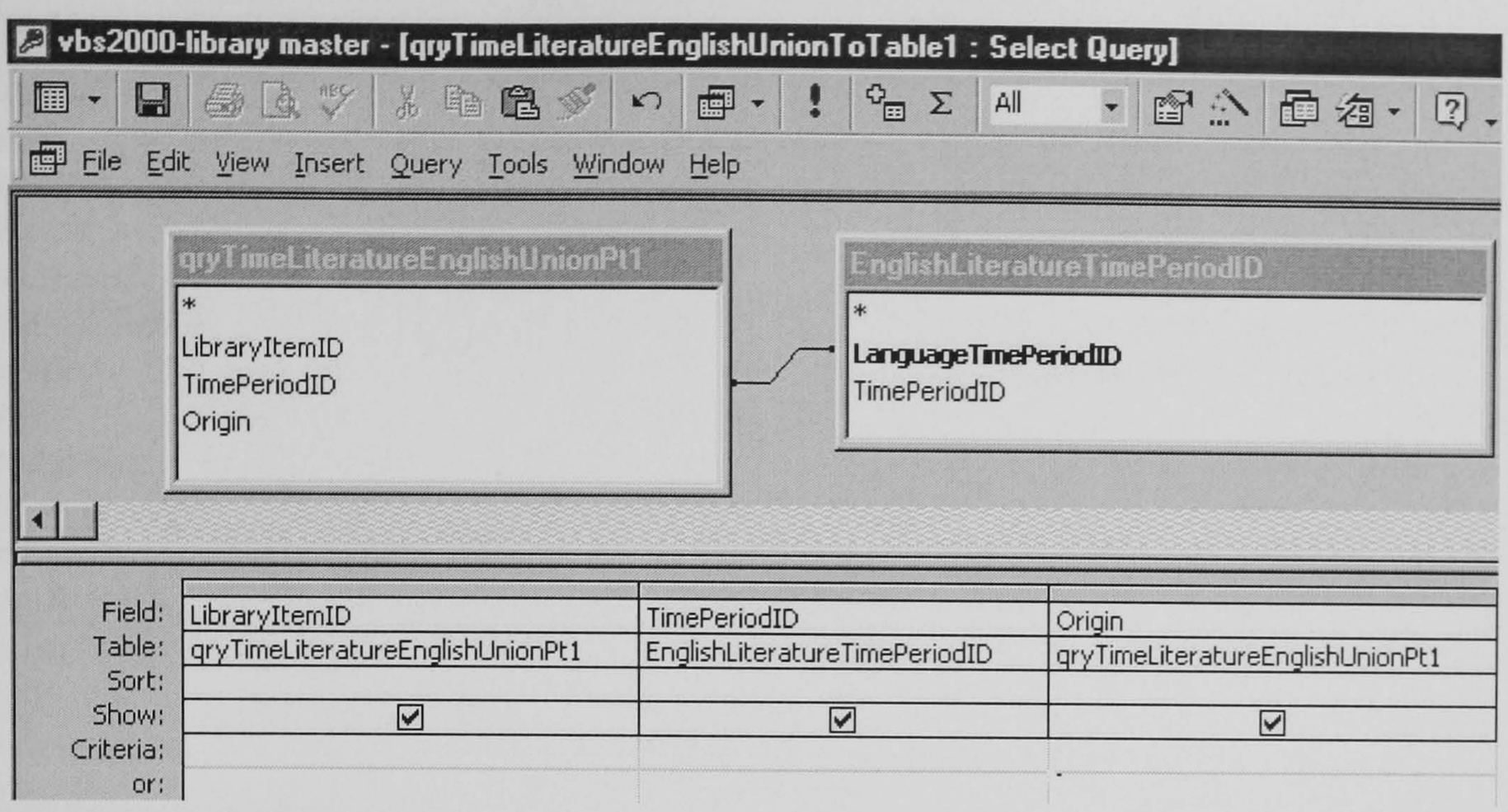


Figure 7.1: Literature by Language Union Query, Joined to Appropriate Translation Table

With the previous language-specific time period notation from Table 7.19 now translated into the appropriate format for populating the Time Facet, as follows:

LibraryItemID	TimePeriodID	Origin
141	T1--0904	Built
380	T1--09034	Built
396	T1--0904	Built
397	T1--0904	Built

Table 7.20: Language Time Period Code Translated to Facet Time Period Code

3. The above query is then combined in a second union query with the citation order 3 query, which did not require translation. This provides a complete result set for each language:

Language	Total Number of Library Items
American	964
English	6972
French	3113
German	1401
Italian	46
Spanish	103
Latin	8
Greek	3
Other English - Old English (Anglo-Saxon)	47
Other French - Catalan, Provencal	1

Other Germanic - Yiddish Frisian Dutch Afrikaans Low German Modern Icelandic Swedish Danish Norwegian New Norwegian	72
Other Greek - Modern Greek	2
Other Italian - Romanian	0
Other Spanish - Portuguese	0
TOTAL	**12,732**

Table 7.21: Library Item Totals for Literature by Language with a Time Facet

Following the three-step process, it was intended to combine the union queries for each of the fourteen languages into a single union query: literature by language, however this proved problematic. Consequently, the three-step process was revisited and all the languages were combined in a union query at the second stage (prior to the addition of citation order 3 query), which was successful. The citation order 3 queries for each language were then successfully combined in a second union query.

In summary, two union queries were used to combine the results of literature by language with a time facet: one combining citation orders 1,2 and 4 and another combining citation order 3.

7.2.5 Combining Sets

Union query 4: bringing it all together

The previous sections in this chapter have described the identification and different means of populating the Time Facet, culminating in four union queries:

Union query	**Number of library items**
Schedule captions and Standard Subdivisions	8959
Literature Generalia,	164
Literature by language citation orders 1, 2 and 4	12,717
Literature by language citation order 3	15
TOTAL	21,855

Table 7.22: Results of the Four Union Queries and Sum Total

These four union queries were combined in another union query, enabling the time facet to be populated ultimately by a single union query. Table 7.22 shows the totals for each individual union query and the sum total of these. The union query retrieved 21,840 (9% of the complete 255,212 library collection), just 15 library items fewer than the individual total 21,855, thus indicating that there was virtually no overlap (i.e. duplicates) amongst the result sets of the four separate union queries.

7.3 The Geographic Area Facet

7.3.1 Facet Identification

Like Time, the geographic or place aspect is one of Ranganathan's five fundamental facets. Collating library items classified as having a geographic aspect enables a user to view items about particular countries in isolation, regardless of subject discipline. Furthermore, a user can then combine this geographic area view with the Dewey view to examine its treatment and distribution across subject disciplines.

7.3.2 Facet Creation

The Geographic Area Facet structure for the Access database table is derived from notations T2--4 to T2--9 *The modern world; extraterrestrial worlds* in Table 2, having a top level arrangement by continent as follows:

T2--4 Europe Western Europe
T2--5 Asia Orient Far East
T2--6 Africa
T2--7 North America
T2--8 South America
T2--9 Other parts of world and extraterrestrial worlds Pacific Ocean islands

Due to the extensiveness of this table, only the top three levels of the classification were utilised with the exception of --41 *British Isles* (owing to the British-bias of the library collection), which was manually extended to provide the full hierarchy.

7.3.3 Sources of DDC Codes

Following examination of the Dewey Decimal Classification in relation to a geographic aspect, three main sources of DDC codes were identified:

- Enumerated Schedule *summary* captions
- Table 2 via the Standard Subdivisions
  - Main Class
  - Division
- Table 2 direct, added under instruction in the Schedules

The subsequent sections discuss, in turn, the identification of library items for each of these sources and their mapping to the facet table, including sections throughout on the combination of query subsets and, finally, the combination of these subsets into a single union query.

7.3.4 Identifying and Mapping Library Items to Facet Table

7.3.4.1 Enumerated Schedule Summary Captions

The three hierarchical levels (Main Class, Division and Section) of the DDC captions were examined manually for the occurrence of particular countries. The first three levels were the focus for this process, owing to the huge number of different countries, which would have been impractical to incorporate into a 'pattern-matching' query. The following summary captions and codes were identified:

061-068 General organizations and museology (in particular counties)
071-079 News media, journalism and publishing (in particular countries)
191-199 Modern western philosophy (in particular counties)
274-279 History of Christianity & Christian church (in particular counties)
314-319 General statistics (in particular counties)
554-559 Earth sciences (in particular counties)
914-919 Geography and travel (in particular counties)
940-999 General history (in particular counties)

The above summary Dewey codes were examined to ascertain at what point and the way in which the geographic area Table 2 component is embedded, leading to a grouping and creation of 5 query sets by virtue of common code 'patterns', as shown below by the examples in Table 7.23.

Dewey code (Example Item)	Library Item Title (Example)	Dewey Summary Caption	Table 2 Code	Table 2 Caption
1) 061, 071 and 191 = Extract from 5th character and preface with "T2--7"				
071.**471**	*Naked news*	Journalism & newspapers in **North America**	T2--7**471**	New York …
071.**3**	*Examining newspapers : what research reveals about America's newspapers*	Journalism & newspapers in **North America**	T2--7**3**	United States
191	*The growth of American thought*	Philosophy of **United States & Canada**	T2--7	North America
061	No titles in library collection	General organizations and museology in **North America**	T2--7	North America
2) 062-067, 072-078 and 192-198 = Extract 3rd and 5th character onwards and preface with "T2--4"				
062	*Directory of UK directories*	General organizations and museology **in British Isles In England**	T2--4**2**	England and Wales
072.81	*The Yorkshire Post : two centuries*	Journalism & newspapers **in British Isles In England**	T2--4**281**	West Yorkshire Metropolitan County
074	*La presse francaise*	Journalism & newspapers in **France & Monaco**	T2--4**4**	France and Monaco
197.**2**	*The hedgehog and the fox : an essay on Tolstoy's view of history*	Philosophy of **former Soviet Union**	T2--47**2**	Western area of Russia
3) 068, 079 and 199 = Extract from 5th character and preface with "T2--"				
068.**4**	*European handbook of organisations*	General organizations and museology in **other geographic areas**	T2--**4**	Europe Western Europe
079.**753**	*Citizen Kay*	News media, journalism and publishing	T2--**753**	District of Columbia (Washington)
199.**438**	*Kotarbiânski's praxiology*	Philosophy in **other geographic areas**	T2--**438**	Poland
4) 274-279, 314-319, 554-559, 914-919 = Extract 3rd and 5th character onwards and preface with "T2--"				
274.**28**	*The early Christian archaeology of North Britain*	History of Christianity in **Europe**	T2--**428**	Northeastern England
314.**585**	*National accounts of the Maltese Islands*	General statistics of **Europe**	T2--**4585**	Malta
559.**89**	*The ice forms*	Earth sciences of **other areas**	T2--**989**	Antarctica
919.**9**	*Out there : the government's secret quest for extraterrestrials*	Geography of & travel in **other areas**	T2--**99**	Extraterrestrial worlds
5) 940-990 = Extract 2nd, 3rd and 5th character onwards and preface with "T2--"				
949.**5**	*The Byzantines*	General history of **other parts of Europe**	T2--**495**	Greece
967.**627**	*Koobi Fora research project*	**Central Africa & offshore islands**	T2--**67627**	Rift Valley Province

Table 7.23: The Five 'Patterns' of Geographic Area in Schedule Captions

Two queries were created for each of these 5 'patterns', identifying and mapping library items to the Geographic Facet. The first query identifies library items having the requisite criteria, including a string expression to extract the embedded geographic code, whilst the second query removes any facets subsequent to the geographic component, signalled by a zero facet indicator. The criteria, expressions and results of the first of these two queries are shown in Table 7.24.

Dewey Codes	Query Criteria	Expression	Library Items
1) 061 071 191	Like "071*" Or Like "191*" Or Like "061*"	Table2CodeX: "T2--7" & Mid$([DeweyCode],5)	33
2) 062-7 072-8 192-8	Like "19[2-8]*" Or Like "06[2- 7]*" Or Like "07[2-8]*"	Table2CodeX: IIf(InStr([DeweyCode],"."),"T2--4" & Mid$([DeweyCode],3,1) & Mid$([DeweyCode],5),"T2--4" & Mid$([DeweyCode],3))	809
3) 068 079 199	Like "068*" Or Like "079*" Or Like "199*"	Table2CodeX: "T2--" & Mid$([DeweyCode],5)	31
4) 274-9 314-9 554-9 914-9	Like "27[4-9]*" Or Like "31[4- 9]*" Or Like "55[4-9]*" Or Like "91[4-9]*"	Table2CodeX: IIf(InStr([DeweyCode],"."),"T2--" & Mid$([DeweyCode],3,1) & Mid$([DeweyCode],5),"T2--" & Mid$([DeweyCode],3))	1066
5) 940- 999	Like "9[4-9]*"	Table2CodeX: IIf(InStr([DeweyCode],"."),"T2--" & Mid$([DeweyCode],2,2) & Mid$([DeweyCode],5),"T2--" & Mid$([DeweyCode],2))	10,056
TOTAL			**11,995**

Table 7.24: Query Criteria, Expressions and Results, based on the Five Schedule 'Patterns' identified

Table 7.24 shows the criteria and expressions utilised in the first query, related to the 'patterns' identified in Table 7.23. Queries 2), 4) and 5) include expressions which remove the characteristic decimal point to enable correct mapping onto the geographic facet. For example, query 4) states that if there is a decimal point within the Dewey code then from the position of the third character take one character (i.e. extract the third digit) and then combine this with the codes from the fifth character onwards, prefaced by "T2—", if there is no decimal point then extract the first three characters of the Dewey code (i.e. in its current form).

It can also be seen from the above expressions in Table 7.24 that an interim Table 2 field (Table2CodeX) was assigned, including codes with subsequent non-geographic

facets, which were later removed by query 2 and the field renamed Table2Code using the following expression:

Table2Code: IIf(InStr([Table2CodeX],"0"),Left$([Table2CodeX],InStr([Table2CodeX],"0")-1),[Table2CodeX])

This expression states that if a code within the Table2CodeX field contains a zero then extract this code up to and including the zero and then take away one character (i.e. the zero) otherwise, if the code does not contain a zero, extract the Table2CodeX in its current form.

These five subsets were then combined by a single union query of library items derived from Schedule summary captions, as follows:

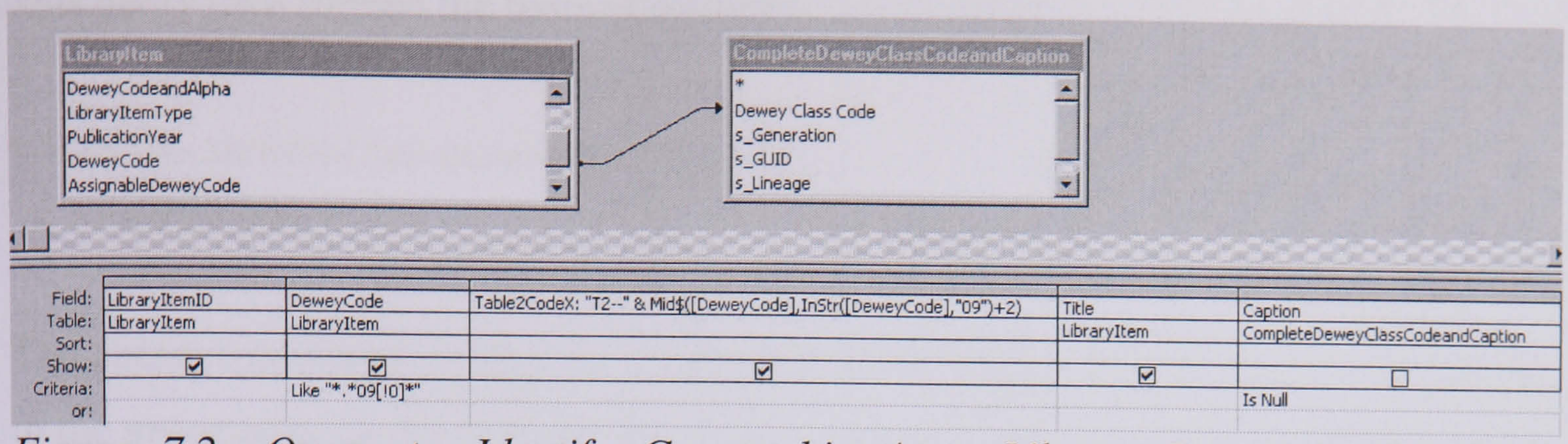
```
SELECT * FROM qryGeographicAreaNorthAmerica071And191And061
UNION
SELECT * FROM qryGeographicArea192To8And062To7And072To8Captions
UNION
SELECT * FROM qryGeographicArea199And068And079OtherAreasCaptions
UNION
SELECT * FROM qryGeographicArea274To9And314To9And554To9And914To9Captions
UNION
SELECT * FROM qryGeographicArea940To999Captions;
```

As shown in Table 7.24, this query retrieved 11,995 library items, constituting 5% of the complete 255,212 item library collection.

7.3.4.2 Table 2 via Standard Subdivisions

In addition to the enumerated geographic components within the Schedules, a Dewey code can be synthesised with a geographic aspect by utilising Table 1 --09 Historical, geographic, persons treatment, followed by, for our purposes, the Table 2 notations -- 4-9 The modern world; extraterrestrial worlds.

A query with the following design was created to identify library items:



Field:	LibraryItemID	DeweyCode	Table2CodeX: "T2--" & Mid$([DeweyCode],InStr([DeweyCode],"09")+2)	Title	Caption
Table:	LibraryItem	LibraryItem		LibraryItem	CompleteDeweyClassCodeandCaption
Sort:					
Show:	☑	☑	☑	☑	☐
Criteria:		Like "*.*09[!0]*"			Is Null
or:					

Figure 7.2: Query to Identify Geographic Area Library Items via Standard Subdivisions

The above query identifies library items with the criteria "*.*09[!0]*"; the exclamation mark preceding the zero, enclosed in square brackets followed by the truncation asterisk means that any character will be found after 09 except zero as 090 would herald a time facet component. The library item table is connected to the facet table of Dewey class codes and captions in a left outer join with the criteria "Is Null" in the Caption field. This means that all enumerated Schedule codes are excluded from the result set, by virtue of having captions, leaving just the required notations synthesised from Tables 1 and 2 and eliminating false drops where such codes are re-used to represent different concepts. The remaining query field contains an expression to extract an interim Table 2 code component (Table2CodeX) by locating the In String facet indicator "09"and extracting a mid string two characters beyond this point, i.e. excluding the "09" component to leave just the geographic code, and prefacing this code with T2--. This query retrieved 14,546 library items, for example:

LibraryItemID	DeweyCode	Table2CodeX	Title
90107	305.560968	T2--68	The rise and fall of the South African peasantry
192963	305.56096897	T2--6897	Structural adjustment and poverty : the case of Malawi
113655	305.560972	T2--72	"We come to object" : the peasants of Morelos and the national state
115313	305.560972	T2--72	Pedro Martâinez : a Mexican peasant and his family
115297	305.560972	T2--72	La Vida : a Puerto Rican family in the culture of poverty-San Juan & New York
72680	305.560972	T2--72	Life in a Mexican village : Tepoztlan restudied
187097	305.560972	T2--72	Five families : Mexican case studies in the culture of poverty
280052	305.56097292	T2--7292	Protecting poor Jamaicans from currency devaluation
68856	305.560973	T2--73	Blue-collar stress
104274	305.560973	T2--73	Race, poverty, and the urban underclass
194902	305.560973	T2--73	Working-class community in industrial America : work, leisure, and struggle in two industrial cities, 1880-1930
189590	305.560973	T2--73	Children and poverty : some sociological and psychological perspectives
203599	305.560973	T2--73	The hidden injuries of class
116360	305.56098	T2--8	Structures of domination and peasant movements in Latin America
279849	305.56098	T2--8	Poverty and income distribution in Latin America : the story of the 1980s
213253	305.56098	T2--8	Indigenous people and poverty in Latin America : an empirical analysis
71009	305.560981	T2--81	The Brazilian peasantry
4179	305.560981	T2--81	The structure of Brazillian development
102783	305.5610941	T2--41	Languages of class : studies in English working class history, 1832-1982
127312	305.5620941	T2--41	The working class in Britain, 1850-1939
131844	305.5620941	T2--41	Fish and chips and the British working class, 1870-1940
114170	305.5620941	T2--41	The remaking of the British working class, 1840-1940
119098	305.5620941	T2--41	Working class cultures in Britain 1890-1960 : gender, class, and ethnicity
27426	305.562095496	T2--5496	Peasants and workers in Nepal

Figure 7.3: Example Standard Subdivision Geographic Area Library Items

This query then formed the basis of a second query with the additional criteria:

Table2Code: IIf(InStr([Table2CodeX],"0"),Left$([Table2CodeX],InStr([Table2CodeX],"0")-1),[Table2CodeX])

The above expression creates the Table2Code field, removing any facets subsequent to the geographic component and enabling a correct match with the Geographic facet table.

There were 10,531 library items matching the geographic facet table, 4015 less than the initial 14,546 result set due to only using the Dewey summaries, with the exception of Great Britain, in the Geographic Facet table.

7.3.4.3 Table 2 via Main Class and Division Standard Subdivisions

The Main Class --09 codes and captions were examined for mention of 'Geographic treatment', which it was found applied to 4 Main Classes: 400-700. A query was constructed with the following criteria to identify library items:

Like "[4-7]09.[4-9]"*

Simultaneously, a query field was created to extract the geographic component, the fifth character onwards:

Table2CodeX: "T2--" & Mid$([DeweyCode],5)

This query identified 559 library items.

The Division Standard Subdivisions entail the --09 Facet Table 1 indicator being split by the decimal point. The following criteria identified library items:

Like "??0.9[4-9]*"

Whilst the following expression extracted the geographic component from the sixth character onwards, retrieving 4971 library items:

Table2CodeX: “T2—“& Mid$([DeweyCode],6)

As in the previous sections, second queries were created for both the Main Class and Division Standard Subdivisions, removing facets subsequent to the geographic component by utilising the expression:

Table2Code: IIf(InStr([Table2CodeX],"0"),Left$([Table2CodeX],InStr([Table2CodeX],"0")-1),[Table2CodeX]).

7.3.4.4 Table 2 added under instruction

A search of the Dewey notes for mention of Table 2 was used to identify those codes having a geographic Table 2 extension, which is added under instruction from the Schedules. This query retrieved 440 codes, which were then examined manually to identify those adding the relevant geographic components, i.e. T2--4-9, for example:

Dewey Class Code	Note
013.9	Add to base number 013.9 notation T2--1-T2--9 from Table 2, e.g., authors resident in Ireland 013.9415
015	Add to base number 015 notation T2--1-T2--9 from Table 2, e.g., works issued in Hong Kong 015.5125 ; then add 0* and to the result add the numbers following 011 in 011.1-011.7, e.g., bibliographies of theses for higher degrees at the university of Hong Kong 015.5125075
020.622	Add to base number 020.622 notation T2--3-T2--9 from Table 2, e.g., Indian Library Association 020.62254, American Society for Information Science 020.62273
020.6232	Add to base number 020.6232 notation T2--3-T2--9 from Table 2, e.g., New England Library Association 020.623274
020.6234	Add to base number 020.6234 notation T2--3-T2--9 from Table 2, e.g., Ontario Library Association 020.6234713
020.624	Add to base number 020.624 notation T2--3-T2--9 from Table 2, e.g., New York Library Club 020.6247471
025.29	Add to base number 025.29 notation T2--1-T2--9 from Table 2, e.g., acquisition of materials from Latin America 025.298

Table 7.25: Example ‘Add to Base Number’ Table 2 Notes in Schedules

The relevant 311 Dewey codes were incorporated into a separate table and matched to the 'Assignable' Dewey code field (library item Dewey codes stripped of their built component). An expression was created which served to then extract just the built component from the un-built 'assignable' code:

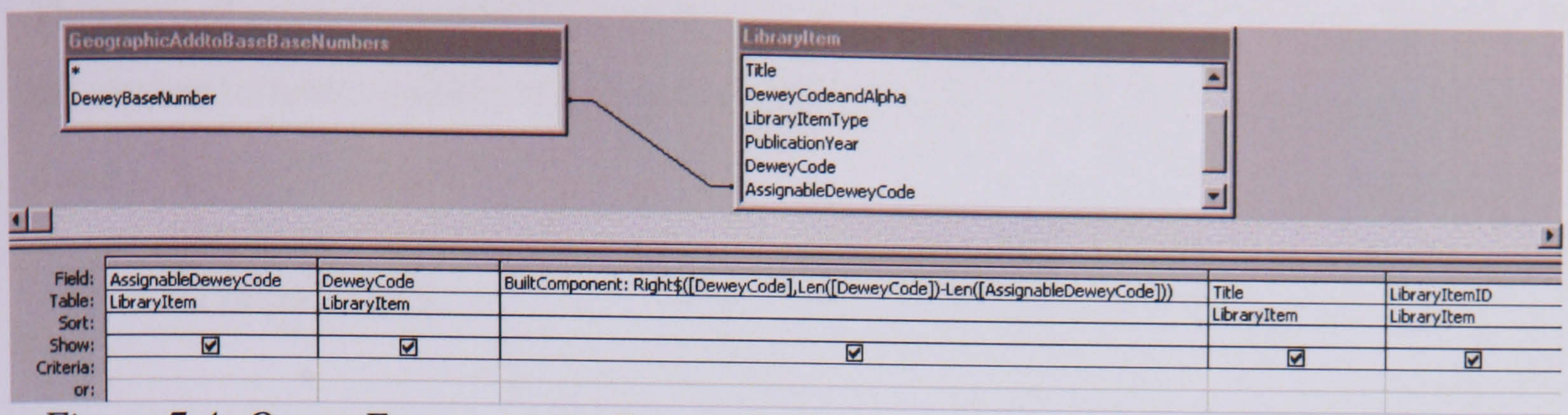


Figure 7.4: Query Expression to Extract Built Geographic Component

AssignableDeweyCode	DeweyCode	BuiltComponent	Title
361.9	361.9**42**	42	Penelope Hall's social services of England and Wales
361.9	361.9**42**	42	English social services : Methods and growth
361.9	361.9**42164**	42164	Organising for social change : a study in the theory and practice of community work
361.9	361.9**4228**	4228	Decisions and resources : report of a seminar held in conjunction with Isle of Wight Social Services Department, 19-21 November 1974
361.9	361.9**42753**	42753	Community work in practice
361.9	361.9**42815**	42815	When social services are local : the Normanton experience
361.9	361.9**5**	5	Social welfare in Asia
361.9	361.9**71**	71	Social welfare in Canada : ideals and realities
361.9	361.9**72983**	72983	A mole cricket called Servol : the early years of an education and community development project in the West Indies
361.9	361.9**72983**	72983	The Servol village : a Caribbean experience in education and community
373	373.**42165**	42165	Charles Edward Brooke School : London Borough of Lambeth Local Education Authority
394.269	394.269**4**0902	**4**0902	City and spectacle in medieval Europe

Table 7.26: Example Schedule Codes, Library Item Codes, Built Components and Titles

This query identified 8091 library items. As can be seen in the last two rows of Table 7.26, the built component is preceded and followed by the decimal point and an additional facet respectively. Consequently, two further queries were created to remove these, enabling a match against the Geographic Facet Table. To remove the decimal point the following expression was created:

Table2CodeX: IIf(InStr([BuiltComponent],"."),"T2--" & Mid$([BuiltComponent],2),"T2--" & [BuiltComponent])

This second query formed the basis of a third query which, as previously, utilised the expression to remove additional facets signaled by zero:

```
Table2Code: IIf(InStr([Table2CodeX],"0"),Left$([Table2CodeX],InStr([Table2CodeX],"0")-
1),[Table2CodeX])
```

However, only 4123 of these 8091 Table 2 Dewey codes matched and hence were used to populate the Geographic Facet Table.

7.3.4.5 Combining subsets

The query result sets of the codes synthesised using Table 2, either via the Standard Subdivisions or directly under Schedule instructions, were combined in a union query, as follows:

```
SELECT * FROM qryGeographicAreaMatchingStandardSubdivisions
UNION
 SELECT * FROM qryGeographicAreaMainClassStandardSubdivisions
UNION
SELECT * FROM qryGeographicAreaDivisionStandardSubdivisions
UNION SELECT * FROM qryGeographicAreaAddToBaseNumber;
```

This query retrieved 18,887 library items.

7.3.5 Combining Sets

To complete population of the geographic facet, both subset union queries (Schedule summaries and the built codes) were combined into one union query, giving a cumulative total of 30,882 library items, constituting 12% of the total 255,212 library collection.

7.4 Conclusion

This chapter has described the derivation and population of the Time and Place Facets, according to the method described in the previous Chapter 5. It was asserted that Time and Place could both be considered fundamental facets, recognised by Ranganathan and having appreciative literary warrant. Despite this, they are afforded a low priority citation order in the DDC and, unless treated as a main subject discipline, are consequently scattered across the classification and library building. It was argued that creation of such facets would allow a user to view library items by a particular time period or geographic place, regardless of subject discipline. The categories and structure for the Time and Place facets was obtained from the DDC Tables 1 (Standard Subdivisions) and 2 respectively. The sources of the DDC codes were similar for each facet. Both utilised the enumerated Schedule captions, although just the Schedule summary captions in the case of the Place/Geographic Area facet, and also the Standard Subdivisions, although the main source of the geographic codes was Table 2, yet added via the Standard Subdivisions. Other sources were time-related codes synthesised within the 800 literature, a worthy yet complex means of facet population, and geographic area components, added under instruction in the Schedules. This range of sources resulted in the creation of numerous queries, the results of which (library item Dewey codes) typically had to have their facet components extracted, re-formulated or translated via transition tables in order to match the Dewey facet table. Despite the difficulty in uniquely and easily identifying time and place components across the DDC, the query results were less prone to 'false drops' than the initial results of the Reference and Persons facets, which had often required exclusion heuristics, as discussed in the previous Chapter 6. As detailed in this chapter, queries based upon positional/citation order criteria proved more reliable, although time-consuming to create. The following chapter evaluates the results of the deconstruction process for all four facets - Reference, Persons, Time and Place – and demonstrates their application within a View-based Searching interface.

CHAPTER 8

FACET ANALYSIS AND APPLICATION AT THE USER INTERFACE

8.1 Introduction

Having derived and populated the four facets with relevant library items, this chapter examines the facet contents in more detail and their subsequent application in a View-based Searching OPAC. Firstly, the contents of the facets are compared statistically and in terms of the origin of their library items. It is shown how library items were retrieved from different origins and can appear under more than one category within a facet. Secondly, it is demonstrated how the populated facets are applied in a View-based interface for end-user searching and browsing and then, thirdly, how these facets can help resolve the problems of scatter, highlighted in previous chapters within this thesis. Finally, consideration is given to both the value and limitations of deconstructing Dewey to derive and populate facets.

8.2 Facet Analysis

The previous chapters 6 and 7 demonstrated in detail how each facet was derived and populated, according to the method described in Chapter 5. The culmination of this work was a union query for each of the four facets, as follows:

Facet	Number of Library Items	Percentage of Total Library Collection
Reference	10, 539	4.1
Persons	11,061	4.3
Time	21,840	8.6
Place	30,882	12.1

Table 8.1: Number of Library Items, Resulting From Each Union Query

Table 8.1 shows the number of library items and their percentage of the total collection retrieved by each union query. However, these union queries combined the

results of several underlying queries, which had identified library items from several different sources. As a result, upon further analysis it can be seen that a library item can be identified by more than one source (e.g. by library item title and by DDC code) and, in some cases, even appear under more than one category within the particular facet; for example, as was discussed in the previous Chapter 6, a library item about both men and women could be included in both facet categories. Consequently, Table 8.1 does not show the true statistic for each facet. The following discussion and table investigates this further and explains how the final figure (number of library items) was achieved for each facet.

Facet	DDC Built	DDC Caption	Item Title	Total
Reference	282	61	285	628
Persons	387	3071	7603	11,061
Time	14,170	7670	--	21,840
Place	18,887	11,995	--	30,882

Table 8.2: Origin and Number of Library Items for Each Facet

Table 8.2 shows the number of library items identified by each particular source: DDC built numbers and DDC captions for the Time and Place facets, together with an additional library Item Title origin for the Reference and Person facets.

These statistics can be more informatively viewed and compared as percentages of each facet, as in Figure 8.1 below.

Figure 8.1: Library Item Origin by Percentage for each Facet

As can be seen from Figure 8.1, library item codes that had be been synthesised (built) in Dewey were the most predominant for the Time and Place facets, suggesting a more efficient re-use of code components for such concepts, as opposed to enumeration (and hence repetition) within the Schedules; similarly, the Reference facet also consists of more built numbers than ones enumerated in the Schedule Captions. In contrast, however, the Person facet was sourced predominantly from library item titles, whose concepts were not represented in their Dewey codes, followed by items with codes enumerated in the Schedules and finally, and atypically, those sourced from built Dewey notations.

The DDC Table of Preference - a guide to determine which Table 1 component to use if more than one Standard Subdivision potentially applies - was examined, yet it seems unlikely that this would have influenced the above statistics. Firstly, library items with more than one Standard Subdivision concept probably occur less frequently, hence the need to choose between such concepts, using the Preference Table, would not arise in the first place. Secondly, even if this situation did apply, the Standard Subdivision 08- code, *history and description with respect to kinds of person*, used in populating the Persons facet is actually quite high in the preference order, taking precedence over facets such as Place and Time; both facets in which built numbers made a considerable contribution. Thus, it seems that library items within the Persons facet either more readily encode their content in the item title; the library item does not cover "the whole, or approximately the whole, subject of the number in the schedules" hence Standard Subdivisions cannot apply (Dewey, 1996, p.3); or the classifier has simply not added the optional Standard Subdivision code. The penultimate and final explanations will be explored further in section 8.5.

Having investigated the composition of each facet in terms of the contribution of each source, its content can be examined in more detail to demonstrate item duplication. As mentioned, a library item can be identified via more than one source, have more than one Library Item ID (i.e. multiple copies available) or appear under more than one facet category (Facet ID), as follows:

qryReferenceItemUnion		
LibraryItemID	**ReferenceType**	**Origin**
737	DEN	Built
737	DEN	Title
757	DEN	Built
757	DEN	Title

Table 8.3: Sample Same Category Library Items, Identified Via Multiple Sources

qryReferenceItem						
Title	**Library ItemType**	**LibraryItem ID**	**ReferenceType**	**DeweyCode**	**AssignableDewey Code**	**Publication Year**
Encyclopedia of sports science	BOK	**278269**	DEN	613.7103	613.7	1997
Encyclopedia of sports science	BOK	**278270**	DEN	613.7103	613.7	1997

Table 8.4: Sample Multiple Library Item Copies under Same Facet Category

qryPersonsItem						
Title	**Library ItemType**	**Library ItemID**	**PersonType**	**DeweyCode**	**Assignable Dewey Code**	**PublicationYear**
Why men hate women	BOK	6551	MEN	305.3	305.3	1993
Why men hate women	BOK	6551	WOM	305.3	305.3	1993

Table 8.5: Sample Library Item Title Assigned to Two Facet Categories

The above Tables 8.3-8.5 demonstrate how the same item can be retrieved via multiple sources/queries (Table 8.3) and how a library item title can appear more than once in a particular facet either due to there being more than one copy, each with its own unique Library Item ID (Table 8.4), or a single copy appearing under more than one facet category (Table 8.5). As mentioned previously, due to this duplication, the statistics retrieved by the Union queries shown in Table 8.2 were not those eventually used at the system interface. These queries required further processing to eliminate duplicate 'origins' and library item copies (as shown in Tables 8.3 and 8.4), although not duplicate facet categories; if appropriate, a library item could appear under more than one.

In order to establish the number of unique library items for each facet category (although, as we have mentioned, an item can appear in more than one), and calculate the total number of unique library items, two further queries for each facet were created. Firstly, each of the four union queries were used to create a table (i.e. a make table query) with the 'origin' field intact. This table was then used to make another table, without the 'origin' field, creating a set of library items with just LibraryItemID and FacetID fields and thus providing accurate statistics for each facet category. In order to calculate the number of unique library items in each facet, a select distinct query can be run against this table to eliminate duplicate LibraryItemIDs. As will be demonstrated in the following section 8.3, this provides a total statistic for each facet. Thus, the above process was followed for each of the four facets and the resultant tables incorporated into the View-based Searching OPAC, with the following results:

Facet	**Unique Library Items**	**Percentage of Total Library Collection**	**Percentage Change from Union Query**	**Sum of Facet Categories**
Reference	485	0.2	- 95.4	528
Persons	8616	3.4	- 22.1	8849
Time	20,609	8.1	- 6.0	21,005
Place	30,761	12.1	- 0.4	30,858

Table 8.6: Statistics for Each Facet

Table 8.6 shows the statistics, derived from the four union queries following additional processing. As can be seen, the figures now represent a smaller proportion of the total library collection than those shown in the previous Table 8.1. However, it can be noted that, as expected, the reduction is most predominant within the Reference and Person facets, primarily due to the incorporation of library item title as a source, which creates duplication. As can be seen above, there was little decrease in the Time and Place facet sets. The following section demonstrates how these sets can be utilised at the OPAC interface.

8.3 Facets at the View-based Searching Interface

The above facet sets (Table 8.6) were incorporated into a View-based Searching (VBS) database (Pollitt et al, 1996; Pollitt et al, forthcoming), underpinned by Microsoft's SQL-Server 2000. It should be noted that, when joined to the library item table within the VBS application (with full library item details) and the Dewey facet tables, the number of library items was slightly reduced for the Time facet (Microsoft Access set = 20,609; VBS set = 20,602) and increased for the Persons facet (Microsoft Access set = 8,614; VBS set = 8,616). The difference was more substantial for the Place facet (Microsoft Access set = 30,761; VBS set = 29,008). The latter difference was due to the fact that the Dewey codes and captions within the Place facet only extend to three levels of the DDC Table 2, with the exception of the British Isles, which, as described in Chapter 7, section 7.4.2, was manually extended to provide the full hierarchy. Thus, if the entire set of the DDC Table 2 had been electronically available, then the facet would have been more extensively populated to include the 1753 (30,761 - 29,008) un-matching library items. The following screen shots show these slightly revised facet sets at the VBS OPAC interface.

The opening screen presents the user with nine available views from which to choose in a direct manipulation driven interface, as opposed to specification led. As shown in Figure 8.2, there are 9 different views onto the library collection; at the top, 6 subject views (colour-coded beige) and, below, 3 non-subject/material type views (colour-coded green). The four derived facets of Time, Place, Person (titled the more explicit Men, Women and Children) and Reference Works, and focus of this current research, are highlighted in red. The remaining five views were 'a given', created for the larger research project, mentioned in Chapter 1, of which this doctoral research is a part.

Each view provides a series of top-level headings with the number of related library items alongside. The first 'All Subjects (Main Dewey Class)' view contains the entire 255,212 collection of library items, classified and browsable by the Dewey Schedules, comparable to the DDC view presented to a user at the physical library shelves. The 'Textiles' and 'Hospitality Management' views both comprise selected relevant captions from the DDC Schedules, which are buried at a deeper hierarchical level and scattered across DDC Main Classes (e.g. Textile Arts = 746, Textile Design and

Manufacture = 677), yet have been promoted to a top-level position to create a dedicated subject specific facet. Creation of such 'bookmark' views were derived from actual bookmarks given to users at the University of Huddersfield to encourage them to look beyond their own subject floor of the library, using DDC notations which may be valuable but are scattered within and across subject floors. These bookmarks were part of the larger research project, subsidiary to the current investigation, and, therefore, will not be discussed further here, but considered under future research in the final chapter. Thus, the generic facets of this current research (Historical Time Period; Men, Women and Children; Places; Reference Works) can be seen highlighted in red at the second and third levels of the search interface, together with the additional given views of 'Type of Library Item' and 'Year of Publication', which have entries for all of the 255,212 library items.

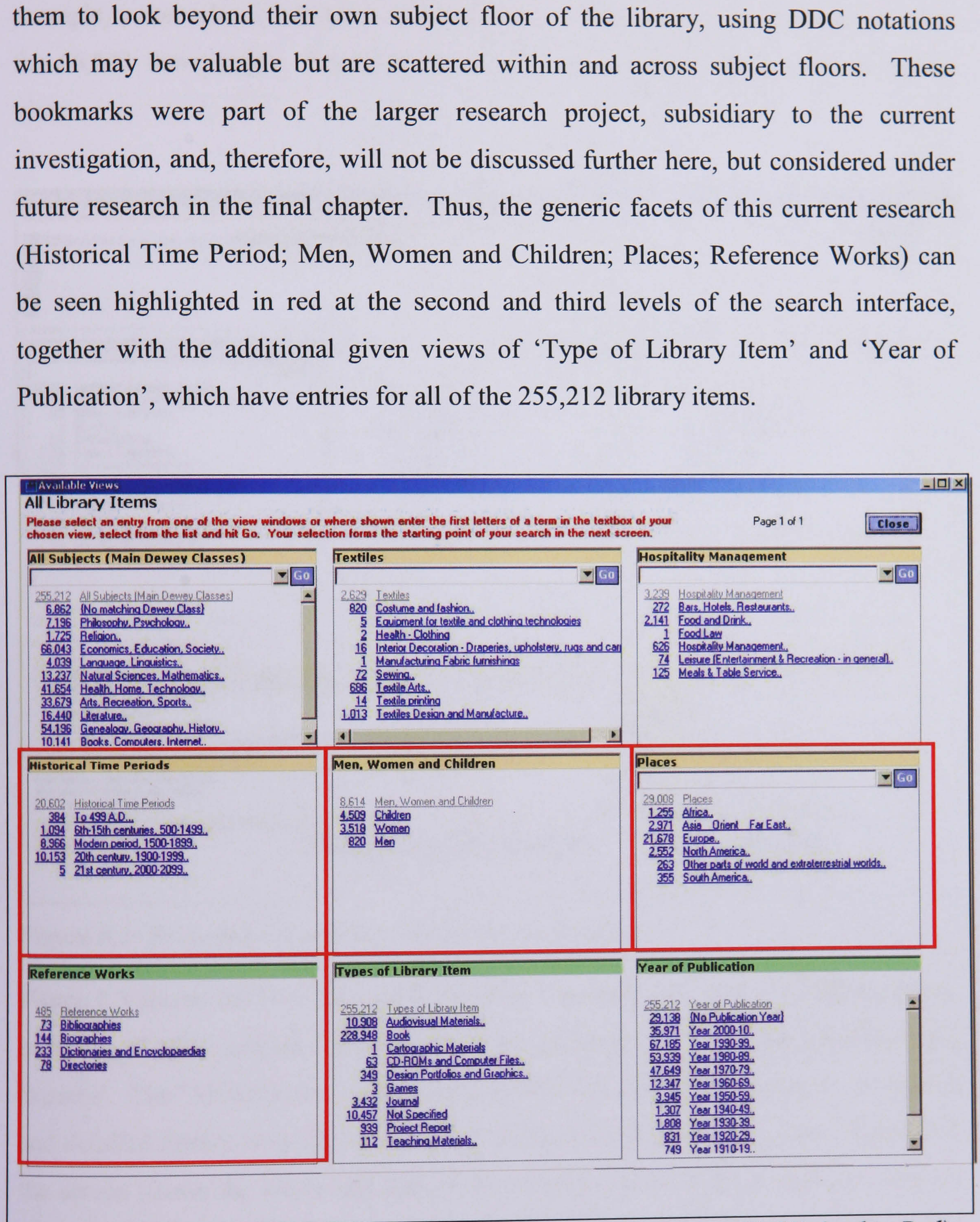


Figure 8.2: Opening Screen of the VBS OPAC (Derived Facets Highlighted in Red)

To begin a search, a user can either enter search terms in any facet's textbox (a keyword search of the Dewey captions/facet categories), or click on any of the categories in a chosen facet for a more browse-based, direct manipulation, search, as

recommended by research discussed in the previous Chapter 3 (e.g. Marchionini, 1995; Hildreth, 1995). A user can therefore initiate a search from a range of different dimensions, courtesy of several facet views within an organised search space. For example, a user who is interested in finding items about higher education might select the 66,043 items under 'Education, Economics and Society' in the 'All Subjects' view:

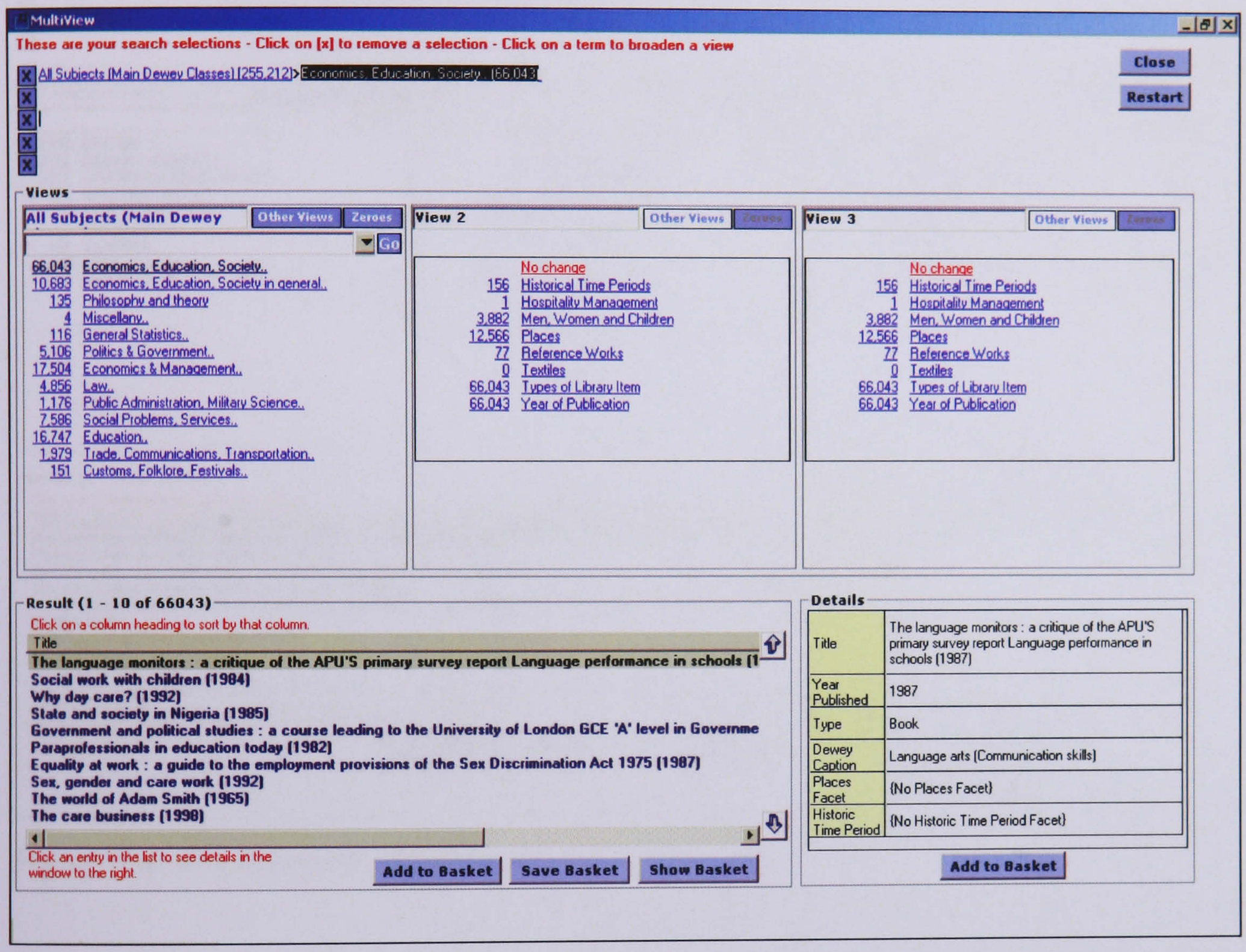


Figure 8.3: Browsing 'Economics, Education and Society'

Figure 8.3 shows the how the search has been narrowed from 255, 212 library items, covering all DDC subject disciplines, to those relating to 'Economics, Education and Society'. The 'All Subjects' view now shows the hierarchy for this caption, with brief and detailed library item title information in the lower half of the screen. The top of the screen shows the views and path of the current search. Here, a user can click on the 'All Subjects Main Dewey Classes [255,212]' to broaden the search to the previous hierarchical level or an entire view can be removed from a search by selecting the cross symbol. If a user wants to incorporate an additional view into the search, the available views and the potential results at the current stage in the search can be seen in the 'View 2' and 'View 3' boxes. For example, at this stage, there are

156 library items relating to a 'Historical Period' *and* 'Economics, Education and Society'. To return to the search for items about higher education, a user might select the 16,747 'Education' items from the 'All Subjects' view:

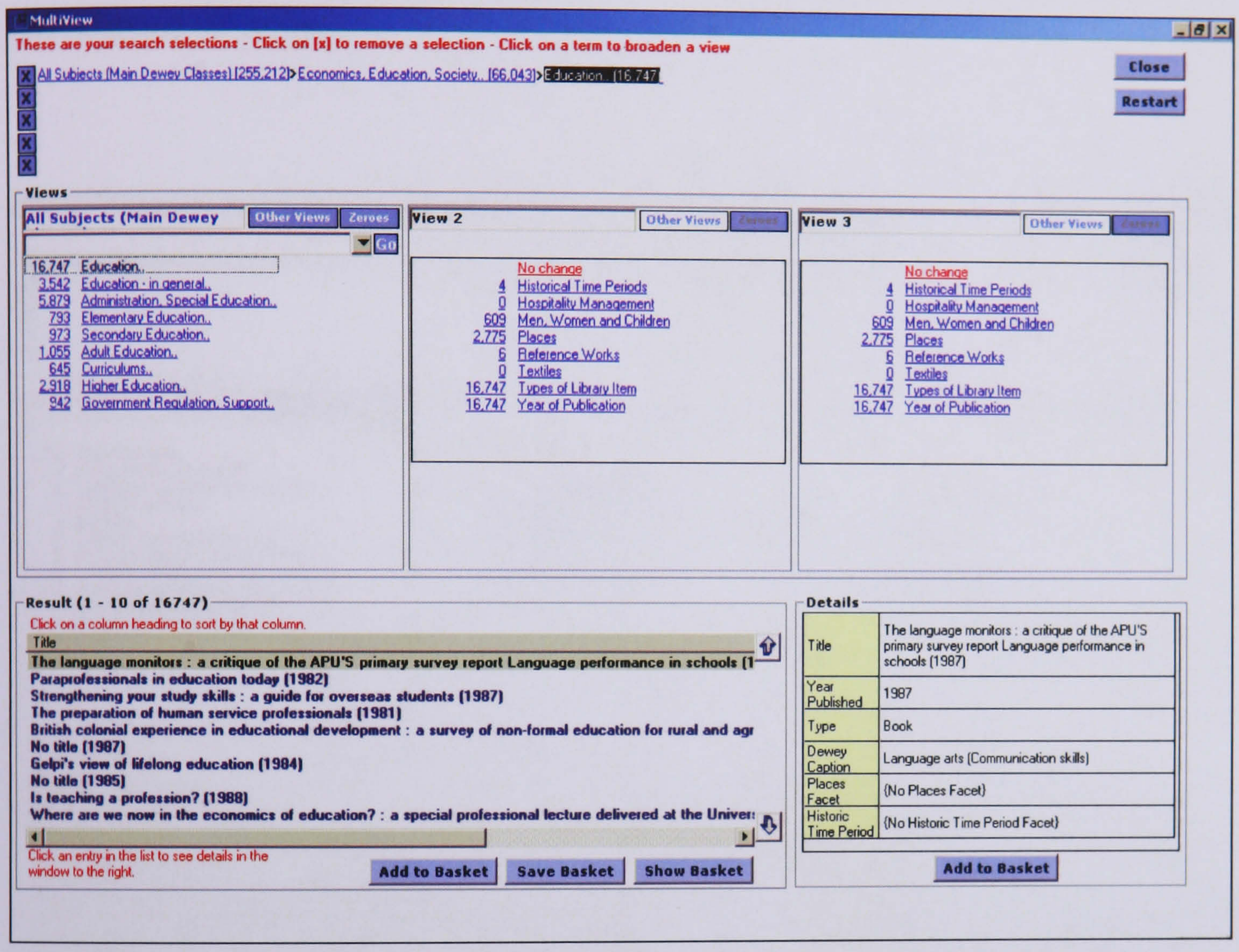


Figure 8.4: Focusing on 'Education'

The search has now been narrowed to the 16, 747 library items about 'Education' and the next level hierarchy is displayed, showing 2,918 library items relating to 'Higher Education'; note how there is also a similar narrowing of the other subject views to the right of the 'All Subjects' view, any of which can be incorporated into the search at anytime. The user focuses on 'Higher Education'. Figure 8.5 shows that there are 2,918 library items about Higher Education and also categories relating to more specific aspects.

The user decides to refine the search further by incorporating two further views: the 643 items within the Places facet and the Year of Publication view, focusing on the category 2000-2010: Although the introduction of additional facets has reduced the number of library items, refined the search, in a sense it has now also expanded the search space, from linear to multi-faceted.

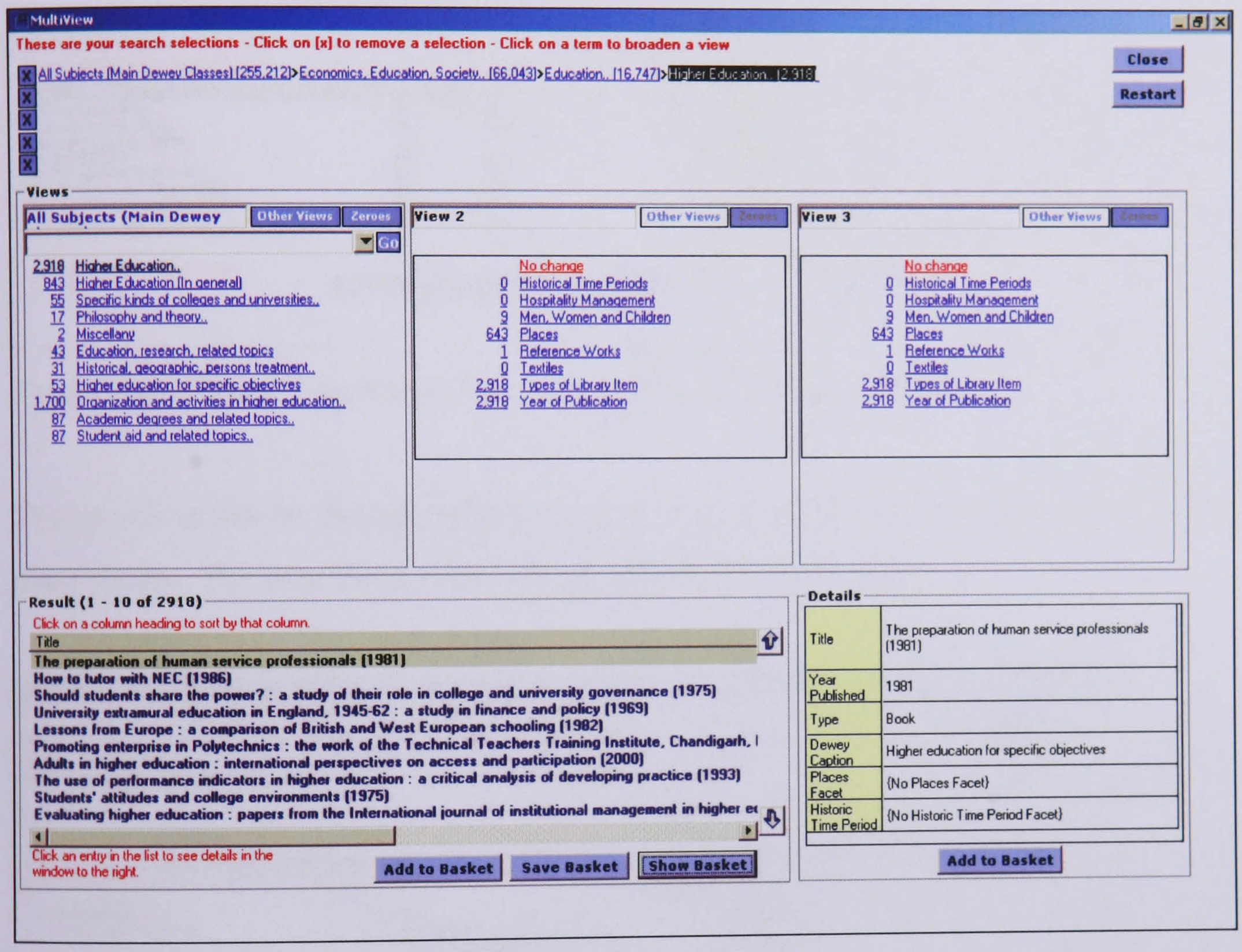


Figure 8.5: Narrowing to 'Higher Education'

Figure 8.6 shows that there are 285 library items about higher education, discussed in relation to a particular geographic area and published between the years 2000-2010.

MultiView

These are your search selections - Click on [x] to remove a selection - Click on a term to broaden a view

Close

Restart

All Subjects (Main Dewey Classes) [3,211] > Economics, Education, Society.. [1,832] > Education.. [564] > Higher Education.. [285]

Places [285]

Year of Publication [643] > Year 2000-10.. [285]

Views

All Subjects (Main Dewey | Other Views | Zeroes | Go

285 Higher Education..
264 Higher Education (In general)
2 Specific kinds of colleges and universities..
17 Organization and activities in higher education..
2 Student aid and related topics..

Places | Other Views | Zeroes | Go

285 Places
5 Africa..
2 Asia Orient Far East..
264 Europe..
14 North America..

Year of Publication | Other Views | Zeroes

285 Year 2000-10..
78 Year 2004
40 Year 2003
91 Year 2002
34 Year 2001
42 Year 2000

Result (1 - 10 of 285)

Click on a column heading to sort by that column.

Title	Year Published	Type
University courses in education open to students from o...	2004	Book
Which degree (2004)	2004	Book
Wearside College (2000)	2000	Book
Myerscough College (2000)	2000	Book
The Ridge (2000)	2000	Book
Prior Pursglove College (2000)	2000	Book
Arnold and Carlton College (2000)	2000	Book
Coalville Technical College (2000)	2000	Book
Henley College Coventry (2000)	2000	Book
Herefordshire College of Technology (2000)	2000	Book

Click an entry in the list to see details in the window to the right.

Add to Basket | Save Basket | Show Basket

Details

Title	University courses in education open to students from overseas (2004)
Year Published	2004
Type	Book
Dewey Caption	Higher Education
Places Facet	British Isles
Historic Time Period	(No Historic Time Period Facet)

Add to Basket

Figure 8.6: Higher Education, Places and Year of Publication Views

The search could be further refined by selecting more specific aspects of any of the three facets. The user focuses on 'Africa' within the Places facet:

MultiView

These are your search selections - Click on [x] to remove a selection - Click on a term to broaden a view

Close

Restart

All Subjects (Main Dewey Classes) [114] > Economics, Education, Society.. [59] > Education.. [26] > Higher Education.. [5]

Places [285] > Africa.. [5]

Year of Publication [21] > Year 2000-10.. [5]

Views

All Subjects (Main Dewey | Other Views | Zeroes | Go

5 Higher Education..
5 Higher Education (In general)

Places | Other Views | Zeroes | Go

5 Africa..
3 Africa (In general)
1 Southern Africa Republic of South Africa..
1 West Africa and offshore islands..

Year of Publication | Other Views | Zeroes

5 Year 2000-10..
1 Year 2004
1 Year 2002
1 Year 2001
2 Year 2000

Result (1 - 5 of 5)

Click on a column heading to sort by that column.

Title
Institutional management in higher education in Western Africa : report of an IIEP sub-regional workshop,
Breaking new grounds for effective university resource management : planning to the year 2000 (2004)
The African Experience with Higher Education (2001)
Education for change in South Africa : reflections on the National Commission for Higher Education (2002)
The African university and its mission : strategies for improving the delivery of higher education institution

Click an entry in the list to see details in the window to the right.

Add to Basket | Save Basket | Show Basket

Details

Title	Breaking new grounds for effective university resource management : planning to the year 2000 (2004)
Year Published	2004
Type	Book
Dewey Caption	Higher Education
Places Facet	Nigeria
Historic Time Period	(No Historic Time Period Facet)

Add to Basket

Figure 8.7: Higher Education, Africa and Publication Year 2000-2010

Figure 8.7 shows 5 library items relating to the above three aspects, which could be refined further by selecting narrower categories within the 'Place' or 'Year of Publication' views.

This section has introduced the basic features of the faceted View-based interface and the nine facets, highlighting the 4 that are the focus of the current doctoral research. Firstly, a user can view the entire library collection within an organised and browsable search space of both subject and material type categories. Secondly, they can initiate a search by recognition and direct manipulation, as evidenced and recommended by research discussed in the previous Chapter 3 (e.g. Marchioni, 1995; Hildreth, 1995), without having to formulate specific search criteria for what may often be vague information needs (Belkin, 1982) or evolving information needs (Bates, 1989; Kuhlthau, 1993, 1999), with users seeking interfaces that encourage exploration not promote specification (Kuhlthau, 1999). A user can then easily narrow or broaden this organised search space at the click of a button rather than having to specify a refined search statement, which often results in a disappointingly large or nil result set (Walker, 1991; Markey, 1989). The section demonstrated an example search interaction, initiated by subject discipline, 'All Subjects (Main Dewey Classes)', as the primary facet. As previously mentioned, browsing this view alone is akin to the linear physical shelf order, with subject discipline as the primary facet; however, by incorporating additional 'hidden' facets such as 'Place', 'Time', and 'Year of Publication' etc, alternative arrangements and a multidimensional search space is revealed. Moreover, a user can also begin a search with non-primary facets that are physically scattered across the classification scheme and library shelves, as demonstrated in the following section.

8.4 Resolution of Scatter

The previous section 8.3 described a search influenced by the predominant classificatory model, and perhaps information need, of arrangement and browsing by subject discipline. However, what of the user who desires an arrangement by embedded, non-primary scattered facets such as time, place etc.? As argued in earlier

chapters, there is the potential for such facets, pre-coordinated by classificatory notation in the physical domain, to be separated, moveable and interchangeable in the electronic domain, expanding browsing from a linear to multi-dimensional platform.

It was highlighted in the previous section how a user could begin a search interaction by selecting categories within any of the nine facets. For example a history student researching developments of the twentieth century may wish to begin a search by selecting this category from the 'Historical Time Periods' facet and then incorporating the 'All Subjects' facet to view the distribution across different subject disciplines, providing further categories to help refine the search:

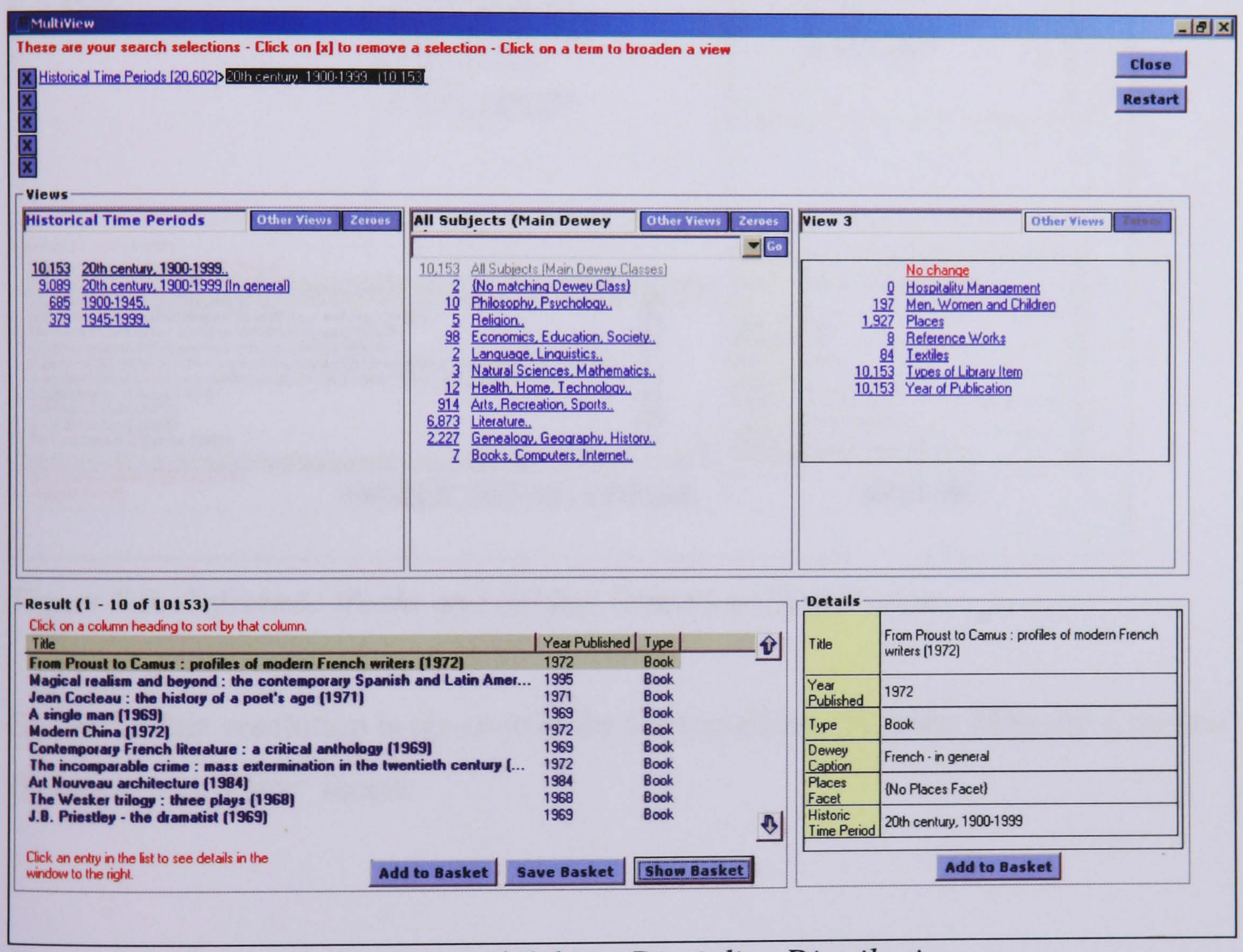


Figure 8.8: Twentieth Century and Subject Discipline Distribution

Figure 8.8 shows how the 10,153 library items about the twentieth century can be viewed both collated and distributed by subject discipline. It can be seen from the above how there is a distribution across all Dewey Main Classes, with the majority of library items (6,873) relating to Literature. A user could therefore achieve a multidisciplinary overview of the twentieth century, embracing all aspects of society at this time, or choose to focus on one particular aspect.

A similar scatter can be observed for the Reference Works facet. Figure 8.9 below illustrates how the 428 different types of reference work are scattered across all the ten Dewey Main Classes. A user can choose to view a colleted list of biographies, for example, and/or examine their related subject area.

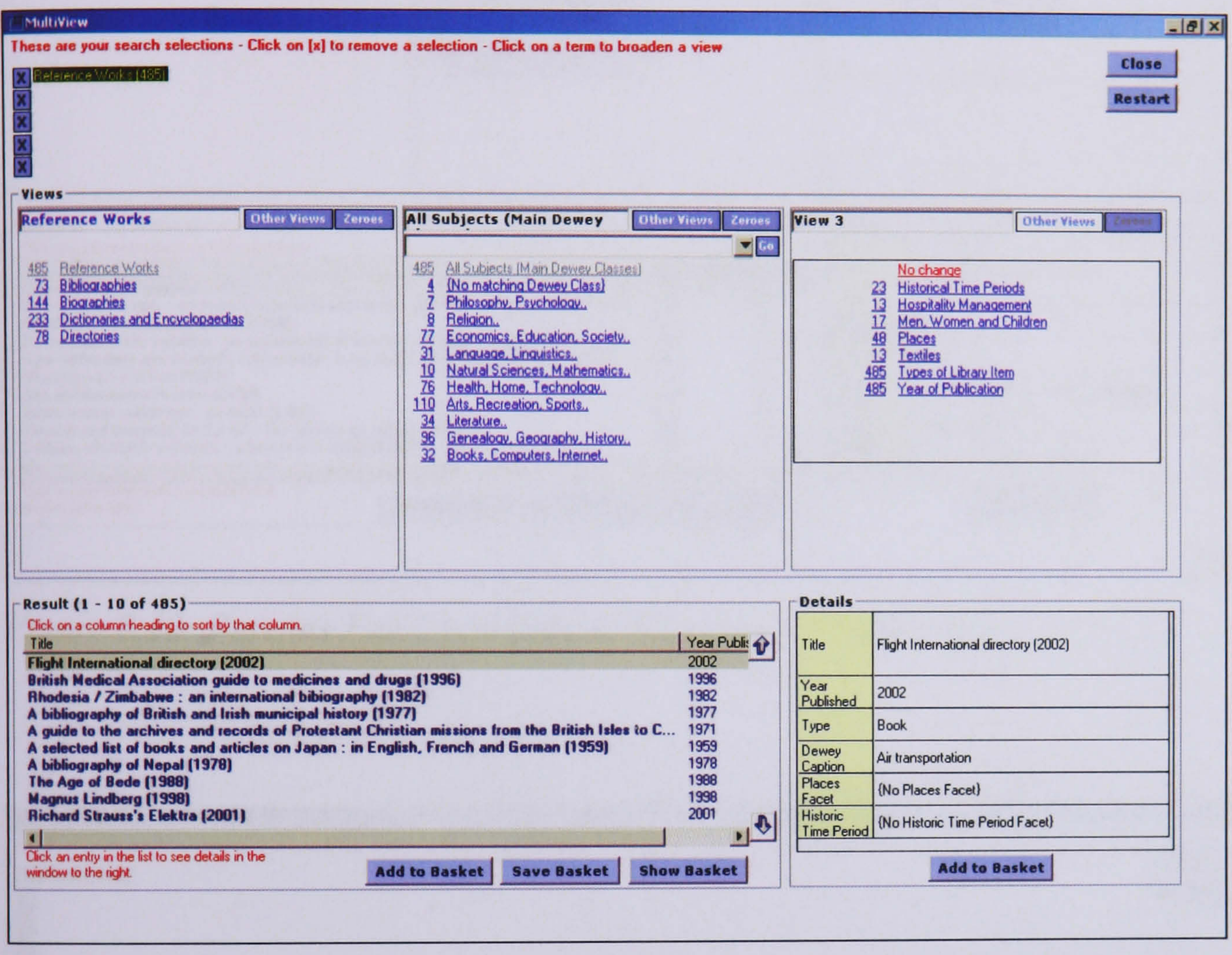


Figure 8.9: Reference Works and Subject Discipline Distribution

Similar scatter resolution is observable for the remaining 'Women, Men and Children' (Persons) and 'Place' facets:

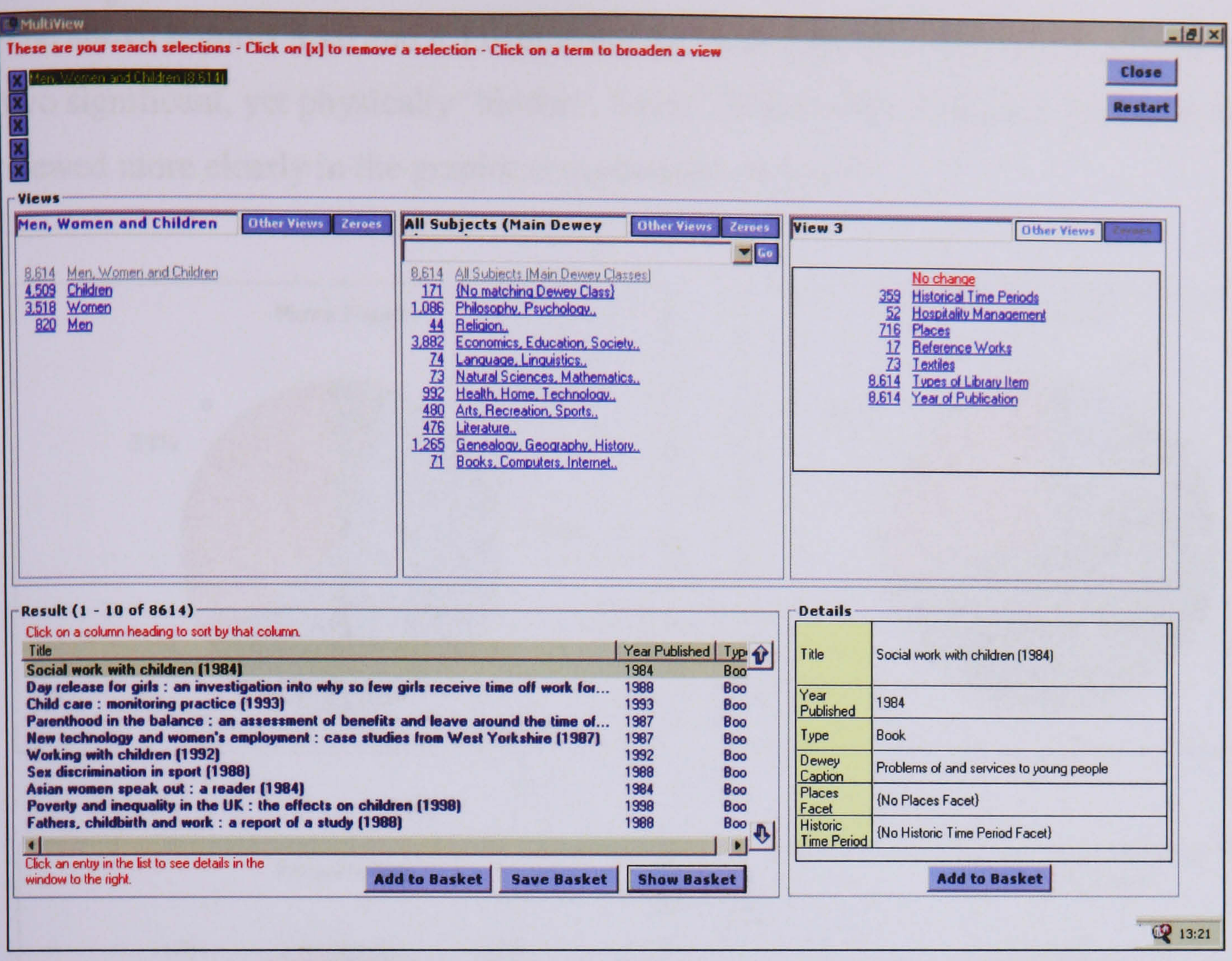


Figure 8.10: Persons Facet and Subject Discipline Distribution

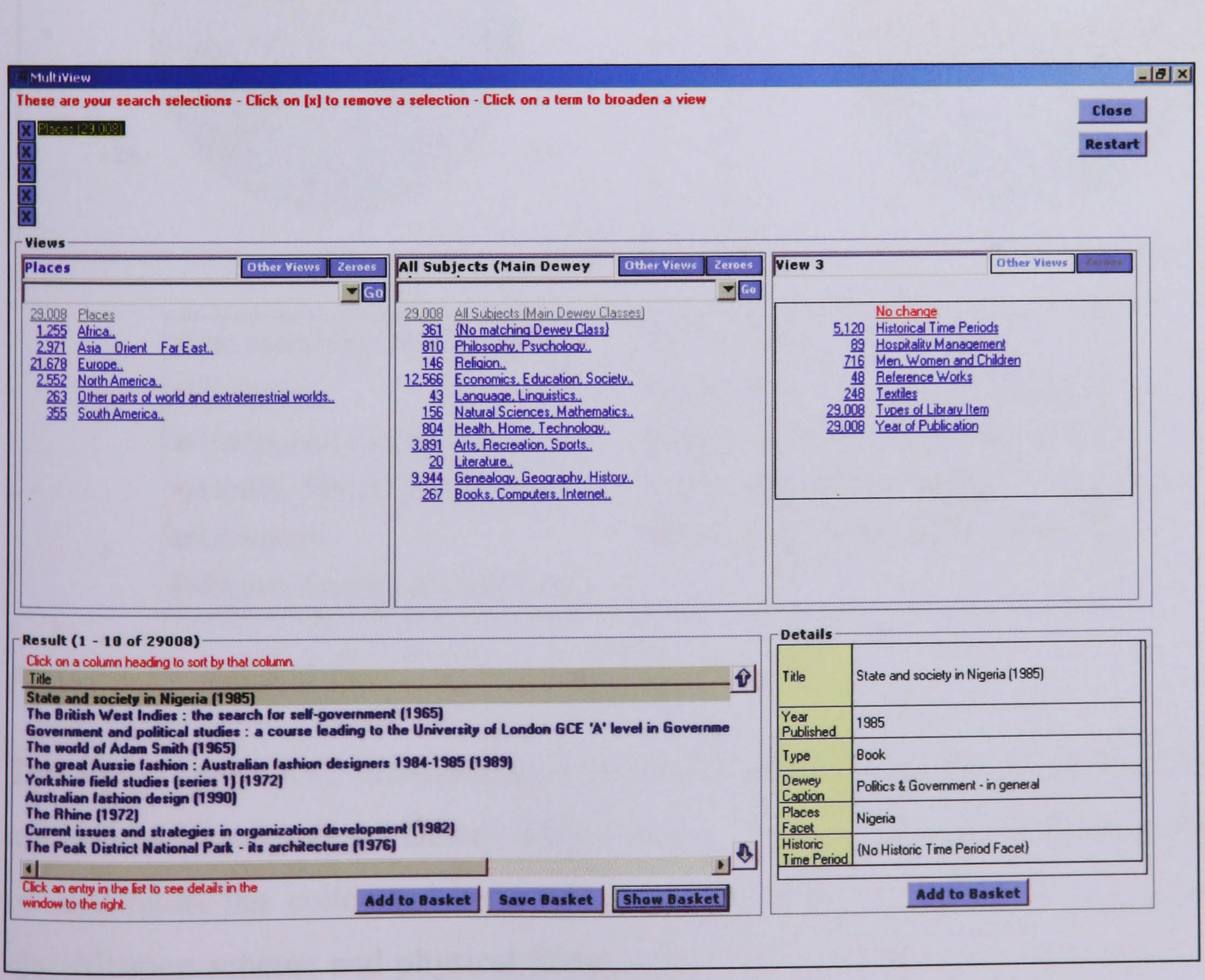


Figure 8.11: Places Facet and Subject Discipline Distribution

The above Figures 8.10 and 8.11 illustrate the simultaneous collation and scatter of two significant, yet physically 'hidden', facets. The scatter of all four facets can be viewed more clearly in the graphical representation below:

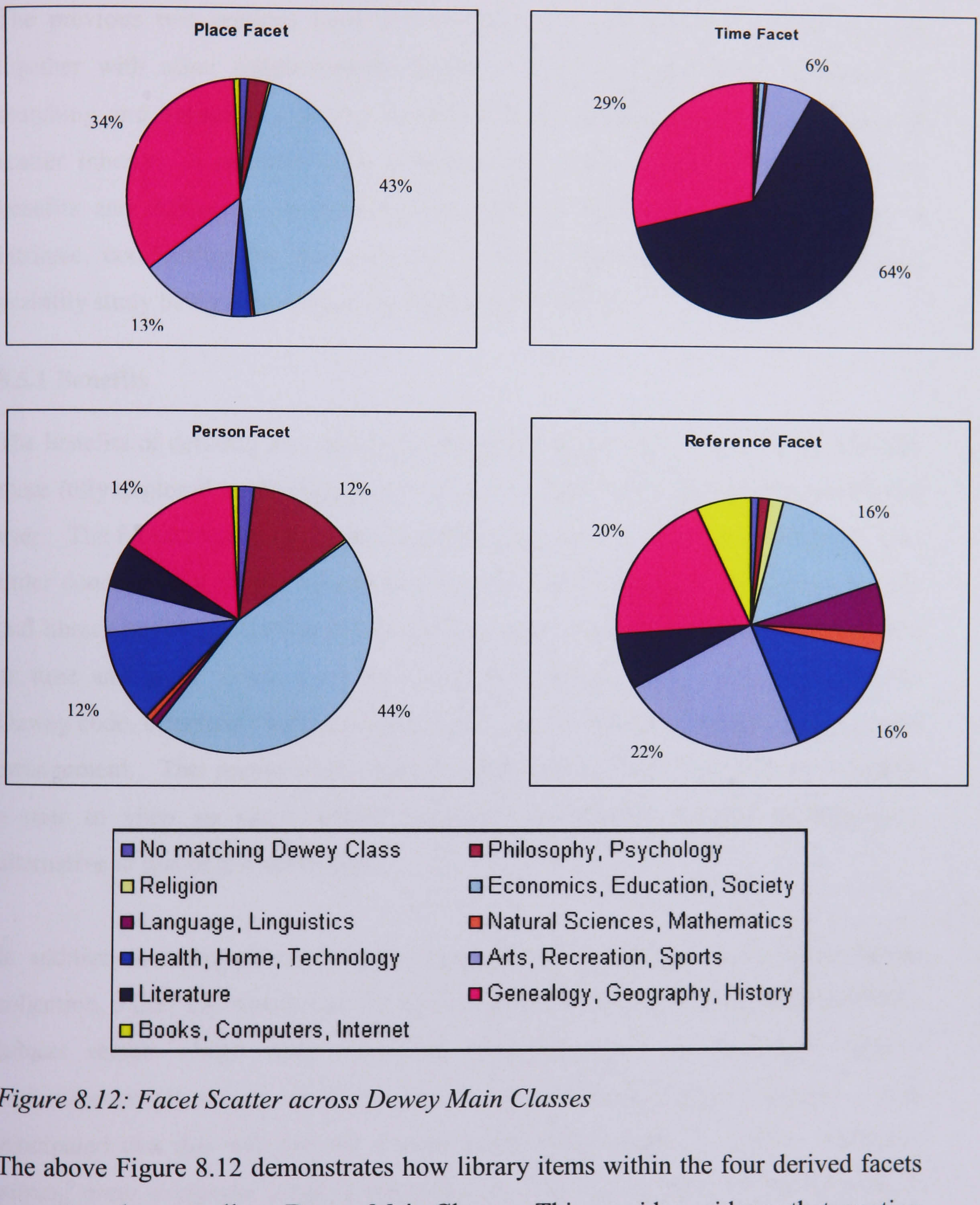


Figure 8.12: Facet Scatter across Dewey Main Classes

The above Figure 8.12 demonstrates how library items within the four derived facets are scattered across all ten Dewey Main Classes. This provides evidence that creation of such facets has collocated items which are otherwise distributed across the entire classification scheme and physical library collection. The value for the user of such collocation by subject facet will be evaluated in the following chapter; however, the

following section considers the intrinsic benefits and difficulties in the facet derivation and deconstruction process.

8.5 Evaluation

The previous two sections have demonstrated how the four facets were applied, together with other supplementary facets, in a View-based OPAC interface for searching and browsing a library collection whilst helping resolve the problem of scatter inherent in physical shelf arrangements. This section considers both the benefits and difficulties in deriving and applying the facets. The evaluation is intrinsic, considering the deconstruction process, with the results of an extrinsic usability study being presented in the following Chapter 9.

8.5.1 Benefits

The benefits of deriving and applying such facets are probably best viewed, and later more fully explored in the subsequent chapter, in terms of the advantages for the end user. The facets themselves collect together library items according to 'hidden', low-order concepts that would otherwise be distributed across the classification scheme and library building. Unless considered a primary concept, fundamental facets such as time and place would have their notational components embedded within the Dewey code, effectively hidden from the user and not brought together within a shelf arrangement. This research has sought to collect by such distributed facets to enable a user to view an entire library collection by several different arrangements, alternative to that of the shelf order.

In addition to being provided with a conceptual overview of an extensive library collection, a user can manipulate the Dewey hierarchies to begin, expand and refine a subject search whilst using the facets in combination, to limit and create a multidimensional search, or alone to view otherwise embedded concepts. It is anticipated that this will provide a more exploratory system for subject searching, helping users overcome some of the problems described in the previous Chapter 3, assisting those with vague information needs and those having difficulty in matching search terms to an unseen system vocabulary in a keyword search. The use of Dewey within this research also has widespread implications. Having the status of the most widely used classification scheme in the world, the facets and underpinning queries

could be applied and tailored to the numerous library collections utilising the scheme. However, despite the research arriving at this point, the following processing caveats should be highlighted.

8.5.2 Difficulties

Three main types of difficulty were experienced in the derivation of the facets, which will be discussed in turn, primarily due to the inconsistency and rigidity of Dewey when attempting to deconstruct:

- Delineation of Facets
- Comprehensiveness of Facets
- False Drops

The delineation of facets within a Dewey code is difficult due to zero being the only, relatively clear, facet indicator, as mentioned in Chapter 2; unlike the UDC, there are no symbols to identify where a facet begins and ends or to signify its type. This makes deconstruction difficult, since the classification is not designed to be taken apart. Furthermore, a Standard Subdivision concept, on which the facets were based, can be added in different (via the optional Table 1 or the same Standard Subdivision codes incorporated into the Schedules) and complex ways, with a Standard Subdivision concept being implied in the Schedule captions, yet having a different notational component. This, again, complicated the deconstruction process, as there was a need to establish the different ways in which Standard Subdivision codes and concepts were synthesised or incorporated within the classification; we cannot take something apart unless we study how it has been put together. Another complication was the partial re-use of Standard Subdivision codes, typically in 'Add to Base Number' instructions or partial incorporation in the enumerated Schedules, which meant that the extracted facet codes required modification to enable a match to the Facet ID code, as demonstrated by this example from the DDC Schedules:

> *708.21-708.28 England*
> *Add to base number 708.2 the numbers following T2--42 in notation T2--421-T2--428 from Table 2, e.g., galleries, museums, private collections in Manchester 708.2733*
>
> OCLC (1998)

The above example illustrates how the Table 2 notation for Manchester (--42733) is re-used and appended to the base number 708.2, but only after removing the digit 4 prefix, signifying Europe.

The comprehensiveness of the facets could also be questioned. As was noted in Chapter 2, a Standard Subdivision concept can be enumerated in the Schedules, added under instruction or, most likely, added on the initiative of the classifier from the optional Table 1, as demonstrated by the high proportion of 'built' numbers within the facets in Figure 8.1. Most frequently, then, the onus is upon the classifier to add a Standard Subdivision component to a library item. Moreover, library policy can often restrict the number of digits within a DDC code, demonstrating the tension between the purposes of classification for online information retrieval and classification for shelf arrangement. For example, at the University of Huddersfield, it is recommended that Dewey codes be no longer than 10 digits. Since the Standard Subdivisions take low priority in the citation order, compared to other subject components, they usually appear as the final component within the Dewey code and thus can be easily excluded to reduce the length of the class number. This would mean that despite having a relevant Standard Subdivision concept, it might not be incorporated within the Dewey code and hence the facet due to this omission. Another practice which could potentially exclude a relevant library item from a facet is if the library item topic is listed within an including note in the Dewey Schedules hence not approximating the whole of the number, as mentioned in the previous Section 8.2, which would preclude addition of a Standard Subdivision code. Ross Trotter exemplifies this difficulty via a personal communication with the author (Trotter, 2000):

> ...a work on rail passenger services in India would indeed be 385.220954 but a work on rail passenger baggage services in India would simply be 385.22 as baggage services is in an Including note.

The above is one of many possible examples of topics currently appearing in "standing room" (Chan et al, 1996, p.28), awaiting a growth in their published literature before meriting their own dedicated Dewey code. A further characteristic of Dewey, which could affect inclusion within a facet, is the Table of Preference citation order. Again, this was mentioned briefly in section 8.2, which explained how a library item with more than one Standard Subdivision concept, e.g. time period and geographic location, could only have one of these facets incorporated into its DDC code, unless instructions permitted otherwise. Thus, for such a library item, place would take precedence over time, since it has a higher ranking in the Table of Preference (Dewey, 1996).

The final type of difficulty was the problem of library item 'false drops', retrieved particularly by the Reference and Person Facet queries. As was discussed, this was due a lack of a uniform notation within Dewey, whereby the same Standard Subdivision code can be re-used to represent different concepts. This led to an analysis of the false drops and the development of heuristics to help counteract this problem. Another cause of a small number of false drops was the use of previous DDC editions for older library items in the catalogue. Such items still remained at a since relocated Dewey notation, which now represents a different topic.

8.6 Conclusion

This chapter has analysed and compared the different origins of the four derived facets (Time, Place, Persons, Reference), demonstrated how these facets were applied at the OPAC search interface for subject browsing and resolving disciplinary scatter, and evaluated the facet derivation process. When populating a facet, library items can be sourced from more than one origin, i.e. enumerated DDC caption, built DDC notation or library item title (Persons and Reference facets only). Analysis of these sources revealed that built notations were most predominant for the Time, Place and Reference facets, suggesting efficient concept re-use; whereas library item title was the most prevalent source for the Person facet and, conversely, built notations least apparent. Following this analysis by origin, four additional queries removed duplicate library items with different origins, leaving a unique set of library items for each facet. These sets were then incorporated into the View-based OPAC. The system interface exploits the facets in multidimensional subject searching and browsing, allowing a user to view previously embedded, unseen and scattered non-primary facets, now collocated into a single facet view. The user is therefore presented with classified arrangements alternative to the physical linear shelf order, which can be further manipulated and refined by several different facet dimensions. Evaluation of the facet derivation process focused on these perceived benefits of this more exploratory-based subject search, whilst recognising the inherent inconsistencies and rigidity of Dewey in derivation of such facets. The forthcoming chapter 9 takes this evaluation further by examining the usability of facet-based searching from an end-user perspective. The usability of the four derived facets for subject searching is tested in a task-based evaluation, combining both quantitative and qualitative measures. Two groups of library users conduct the same tasks, each group using a different interface: the direct manipulation View-based OPAC and its Dewey facets in comparison to the current specification led (form fill-in) OPAC. The following chapter investigates the response of the user: how does facet-based searching compare to familiar form fill-in approaches in terms of user perception and task completion? Which best facilitates the complex activity of subject searching?

CHAPTER 9

PRELIMINARY USABILITY STUDY OF FACET-BASED SEARCHING

9.1 Introduction

An important aspect of any information retrieval system, over and above the search algorithms, relates to how well a system will be received and used by a user population. This question belongs in the domain of human–computer interaction and usability engineering. According to Preece et al (2002), usability involves three elements: effectiveness (the extent to which a system retrieves accurate objects/output), efficiency (related to the degree of resources used in order to produce that output) and satisfaction (a concept which relates to subjective opinions about the acceptability of computer products). A system which is high in effectiveness and efficiency but low in satisfaction would not be acceptable to a user community. The purpose of the experiment described in this chapter was to compare the perceived usability of a facet-based VBS interface with a traditional form fill-in OPAC system.

Analysis of the usability of facet-based searching has already been attempted in a paper comparing use of a hierarchical faceted interface with a conventional keyword search interface for searching and browsing image collections (Yee et al, 2003). These authors concluded that their Flamenco interface, illustrated in the previous Chapter 3, was a successful approach, with 90% of 32 participants preferring this to the keyword-based search. The current experiment provides further evidence to test the proposition that facet-based searching offers a viable solution to complex subject searches, such as those discussed in Chapters 2 and 3.

9.2 Rationale and Usability Testing

Usability testing is an important evaluation technique within the field of human-computer interaction, assessing whether a system meets its purpose by asking users to perform typical set tasks that relate to the system design (Preece et al, 2002). The performance of users is evaluated by way of metrics such as task timings and number of errors; users are often observed on video or by way of data recorded from system event logs. This data is then often supplemented by eliciting users' opinions by tools such as interviews and questionnaires. Usability studies are typically held in controlled settings, although variables are not controlled; involve small numbers of participants; and emphasis is on acquiring quantitative metrics, with average statistics calculated, together with participant opinions and evidence to illustrate problem areas (Preece et al, 2002).

This preliminary usability study intends to assess the viability of facet-based searching for complex subject searching activities, in comparison to a conventional keyword driven OPAC. Usability testing is clearly an appropriate means of evaluating facet-based searching (the View-based OPAC); typical information retrieval tasks could be specifically designed to test the use and viability of each of the four derived facets within the context of the facet-based interface, using the familiar form fill-in OPAC as a comparative benchmark. Event log data could be captured, enabling the subject search terms to be recorded and navigation paths tracked (the latter particularly for the View-based OPAC) and incidences of empty result sets identified; a typical problem in OPAC subject searching, as was detailed in Chapter 3. Having used the system, inclusion of an attitudinal questionnaire would also then obtain users' opinions as to what extent the users felt the system helped them complete the set tasks. A similar task-based method was utilised in the aforementioned study by Yee et al (2003). One difference, however, was that the Yee et al study was based upon a within subjects design, in contrast to our between subjects design. The main rationale for this was that our users were already familiar with the form fill-in OPAC (unlike participants in Yee's study, who had not used the baseline system previously) and the form fill-in OPAC was accessing a larger, more current database of library items. Other areas of this current usability study maintain

a certain parallel with Yee et al's research (e.g. task-based, task timings, eliciting participant opinions), in order that some comparisons might be drawn.

9.3 Method

9.3.1 Participants

The 34 participants were all students at the University of Huddersfield and were recruited either in response to a lecturer request for participation or a personal approach from the author or library staff whilst the student was working in the library. Participants were offered a £10 library photocopy card or print credits as a reward for participating. As a library OPAC is a walk-up-and-use system, aimed at all types and levels of user, there was no participant control for age, sex, year and level of study or subject area. All participants had previously used the current library OPAC (IPAC system).

9.3.2 Materials

Systems

A View-based prototype system (VBS) of that described in Chapter 8 was developed, which was to be compared to the current University of Huddersfield library system (IPAC) interface:

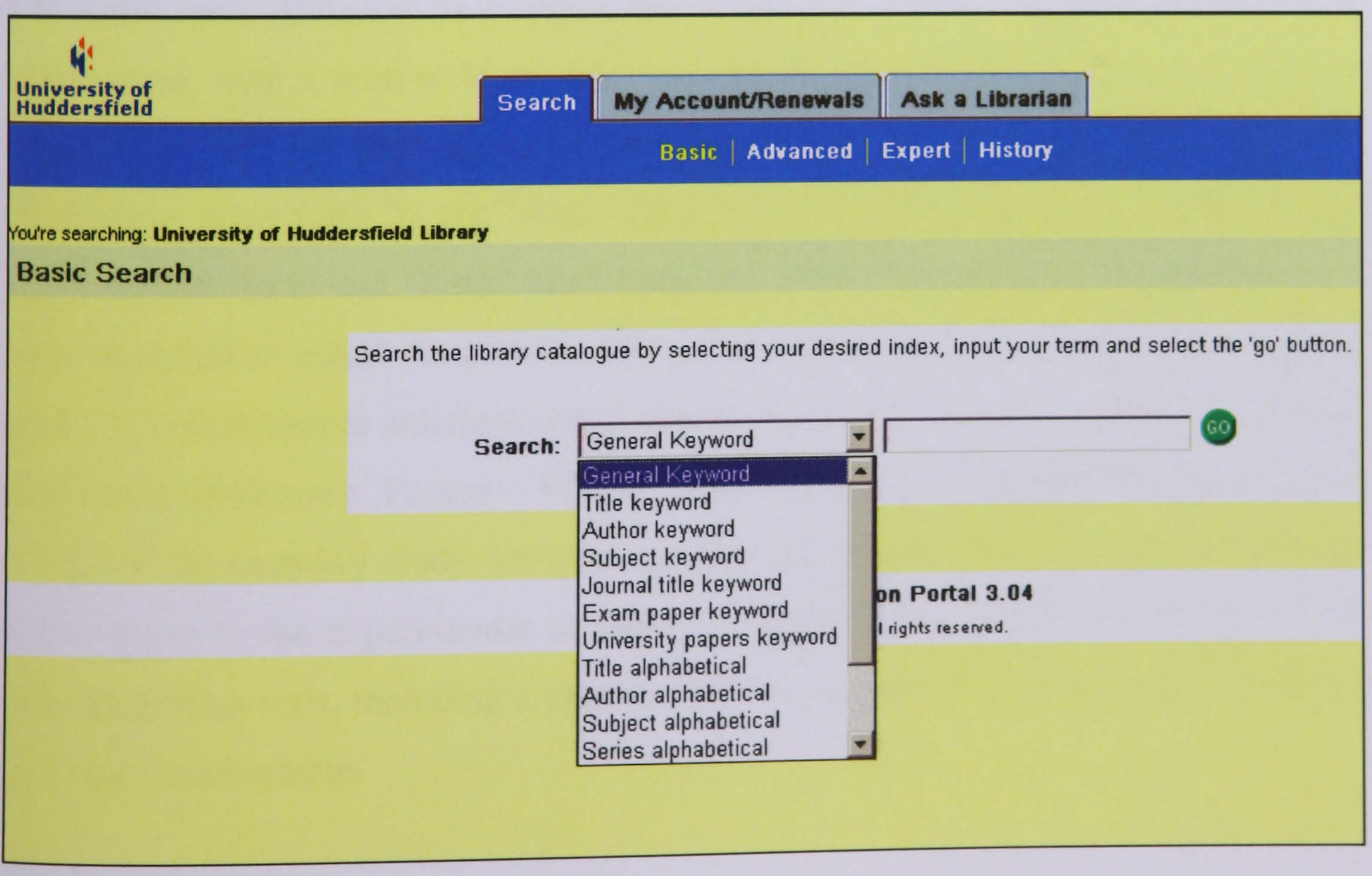


Figure 9.1: IPAC Search Interface

Figure 9.1 shows the opening search screen, the search box, in which users enter keywords, and how a search can be restricted by selection of various indexes.

Tasks

The tasks were all subject searches (i.e. not known-item title or author searches) and were designed to include more than one subject component (facet). Each participant completed a practice task, followed by four main tasks, as follows:

- *Practice task*
  Find 3 items about the architecture of public buildings in Britain.

- *Task 1*
  Find 3 items for a literature review about higher education institutions in West Yorkshire.

- *Task 2*
  Find 3 biographical works about 3 different classical composers.

- *Task 3*
  Find 3 works of English poetry by 3 different 18th century poets.

- *Task 4*
  Find 3 items discussing the work of English female novelists in the 19th century.

The tasks were designed to examine the usage of the 'All Subjects (Dewey Main Class) view, with a need to browse beyond the top level of this hierarchical view, in conjunction with the four derived facets: the 'Places' (practice task and task 1), the 'Reference' facet (task 2), the 'Time' facet (task 3), and the 'Person – Women, Men and Children' facet and 'Time' facet (task 4). Thus, the practice task and tasks 1-3 were designed to entail the combination of two facets and task 4, a combination of three. It was therefore anticipated that users would have the opportunity to utilise all four facets (Reference, Person – Women, Men, Children, Time and Place) during the course of the usability study, completing their subsequent system evaluation on this basis. Prior to the experimental sessions, the tasks were checked for clarity in two individual pilot tests, including a check on their suitability for use on either the IPAC or View-based system.

Attitude Questionnaire

At the end of the session, participants completed an attitude questionnaire, designed and validated in previous usability research (Brooke, 1986). Prior to the experimental sessions, the questionnaire was pilot-tested for comprehension by several University of Huddersfield library catalogue users, who were asked to complete a questionnaire and highlight any confusing questions. The questionnaire contains ten questions, requesting a response on a five-point Likert scale. An additional question, asking participants for comments about the library system they had used, was included at the end of the questionnaire to obtain qualitative data. A copy of the questionnaire can be found in Appendix 2.

9.3.3 Research Design and Procedure

The experiment incorporated a between subjects research design in which participants completed the tasks set using either the VBS or the IPAC system. Participants were allocated to a VBS or IPAC session randomly. The experiment was blind with regard to the actual tasks used in the session, although people were told of the purpose of the investigation and then informed consent was gained.

A total of 7 sessions were conducted, 3 VBS sessions and 4 IPAC. There was no time limit on each session. Upon arrival, participants were given a brief spoken introduction to the procedure of the experiment and the system they would be using and, in the case of the VBS system, a short system demonstration; a demonstration was not given to the IPAC participants, as they had all used the system previously. In the case of the VBS system, participants were also given a few minutes to familiarise themselves with the system and find items of interest. Participants were informed that the session aimed to evaluate the system, not their performance: we were not testing them.

For each system, participants were then instructed to begin the practice task, which was displayed by way of a data projector. Participants were informed that questions could be asked during this practice session but not when the tasks began proper. Following completion of the practice task, tasks 1-4 were completed in turn, waiting for all participants to finish each task before the next was displayed and begun. System transaction logs, monitoring keystrokes and task timings for each participant, were captured throughout each session and each participant was provided with a floppy disk to save their search task results (i.e. 3 selected items). Following the four tasks, participants were asked to complete the attitude questionnaire with reference to the extent to which they felt the system had helped their completion of the tasks.

9.4 Results and Analysis

The data from the questionnaires were collected and each questionnaire was scored. The scoring technique used involved separating positive/negative statements and the scores for each negative statement were reversed as shown in Table 9.1. This produces a score for each individual statement and also allows the computation of a System Usability Metric, which is a number in the range of 10-50, with 10 representing poor usability, 30 neutral and 50 indicating a very usable system.

	Actual Rating				
Valence	**1**	**2**	**3**	**4**	**5**
Positive	1	2	3	4	5
Negative	5	4	3	2	1

Table 9.1: Scoring System
Note: This table shows the scoring resulting from participants' actual ratings.

In addition to the overall metric, providing indications of participant attitude to interface type, data relating to a number of sub-themes were also computed, according to scheme shown in Table 9.2.

Sub-theme	Question Number	Question Statement
Learnability	7	I would imagine that most people would learn to use this system very quickly.
	10	I needed to learn a lot of things before I could get going with this system.
Ease of Use	2	I found the system unnecessarily complex.
	3	I thought the system was easy to use.
	8	I found the system very cumbersome to use
Functional Integration	5	I found the various functions in this system were well integrated.
	6	I thought there was too much inconsistency in this system.
Confidence in System	4	I think that I would need the support of a technical person to be able to use this system.
	9	I felt very confident using the system.
User Satisfaction	1	I think that I would like to use this system frequently.

Table 9.2: Sub-themes and Related Questionnaire Statements

From the original source document, it was suggested that item 8 (I found the system very cumbersome to use) was part of the user satisfaction construct. It was felt, however, that in this particular situation the statement was a reflection of ease of use and this was how the data was treated in this evaluation.

As shown in Table 9.2, there are 5 sub-constructs which contribute to a person's overall system usability metric (SUM). These constructs are Learnability, Functional Integration, Confidence in System, all captured by 2 questionnaire statements, and Ease of Use and User Satisfaction, captured by 3 and 1 questionnaire statements respectively.

The raw data for the experiment is provided for reference purposes in Appendix 3.

To commence the data analysis proper, descriptive statistics were computed for the overall SUM score and each of the 5 underlying constructs. The mean was then calculated for each of the 5 constructs for both systems, as shown below in Figure 9.2.

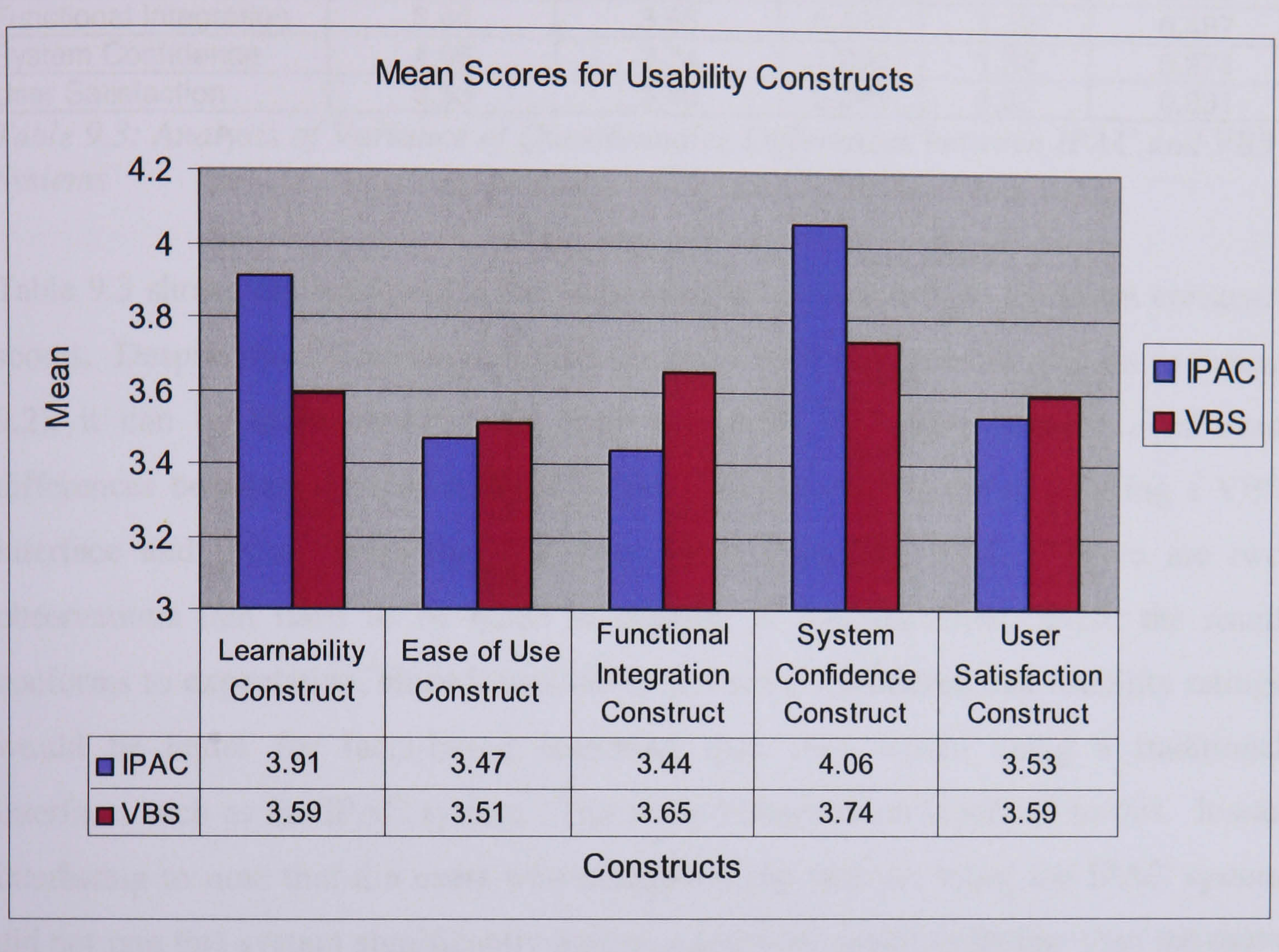


Figure 9.2: Comparison of Mean Construct Scores for IPAC and VBS Systems

As can be seen in Figure 9.2, scores were similar for both systems yet the VBS system was rated more highly for 3 out of the 5 constructs (Ease of Use, Functional Integration and User Satisfaction. In contrast, the IPAC system scored more highly for Learnability and System Confidence. As will become evident, this reflects findings in other areas of the current study, namely that as users became more familiar with the VBS system their task timings improved significantly and qualitative participant comments also emphasised that users felt that they became more comfortable with the system as experience was gained. A comparison of the mean SUM scores (System Usability Metric) across the 5 constructs is, however, weighted in favour of the IPAC system (IPAC = 36.76, VBS = 36.06). However, as can be seen from Table 9.3, these differences were not statistically significant.

Measure	IPAC Mean Score	VBS Mean Score	F	df	p
System Usability	36.76	36.06	0.076	1,32	0.784
Learnability	3.91	3.59	0.797	1,32	0.379
Ease of Use	3.47	3.51	0.017	1,32	0.896
Functional Integration	3.44	3.65	0.494	1,32	0.487
System Confidence	4.06	3.74	1.238	1,32	0.274
User Satisfaction	3.53	3.59	0.046	1,32	0.831

Table 9.3: Analysis of Variance of Questionnaire Differences between IPAC and VBS Systems

Table 9.3 shows the results of a one-way analysis of variance on the mean construct scores. Despite the difference in means (as previously discussed in relation to Figure 9.2), it can be quite clearly seen from this table that there was no significant differences between the responses of participants completing the tasks using a VBS interface and those completing the same tasks using the IPAC. There are two observations that need to be made in relation to this outcome. First, the result conforms to expectation, since it was never seriously considered that usability ratings would be better for facet-based searching than they would using a traditional interface, such as the IPAC system. The second observation is related to this. It was interesting to note that the users who completed the task set using the IPAC system did not rate this system significantly higher in terms of usability factors than the users of the VBS system. This result was obtained despite the fact that both groups of participants were extremely familiar with the IPAC interface and search style/procedure. Consequently, this could provide evidence of the value of a VBS type interface and a facet-based search approach for the naive user.

Thus, the analysis of subjective self-report data, obtained from the usability evaluation questionnaire indicates that there was no significant difference between the perceived usability of either system. This finding was obtained even when a more detailed analysis was performed of the constructs underlying the concept of system usability. Secondary analysis was performed of system usability using more objective data. This data included task timings, system event logs, and number and quality of hits returned.

Task timings were obtained from data logs, created at the time of the experimental session. These logs covered a variety of data but included task start and end times, which enabled the computation of a session time for each subject. A mean time was calculated for each of the 34 participants across the 4 tasks (see Appendix 3 for participant task timings). These data were submitted to a one-way analysis of variance and the results were found to be significant ($F = 5.626$, df 1,32, $p=0.024$, significant). This, therefore, demonstrates a significantly shorter average task time for participants using the View-based system, in comparison to the IPAC system. Thus, it would appear from the results of this particular experiment that View-based searching significantly improves task accomplishment in a library information retrieval situation, compared with the traditional IPAC system. This can be seen most clearly in the plot of specific task timings shown below in Figure 9.3.

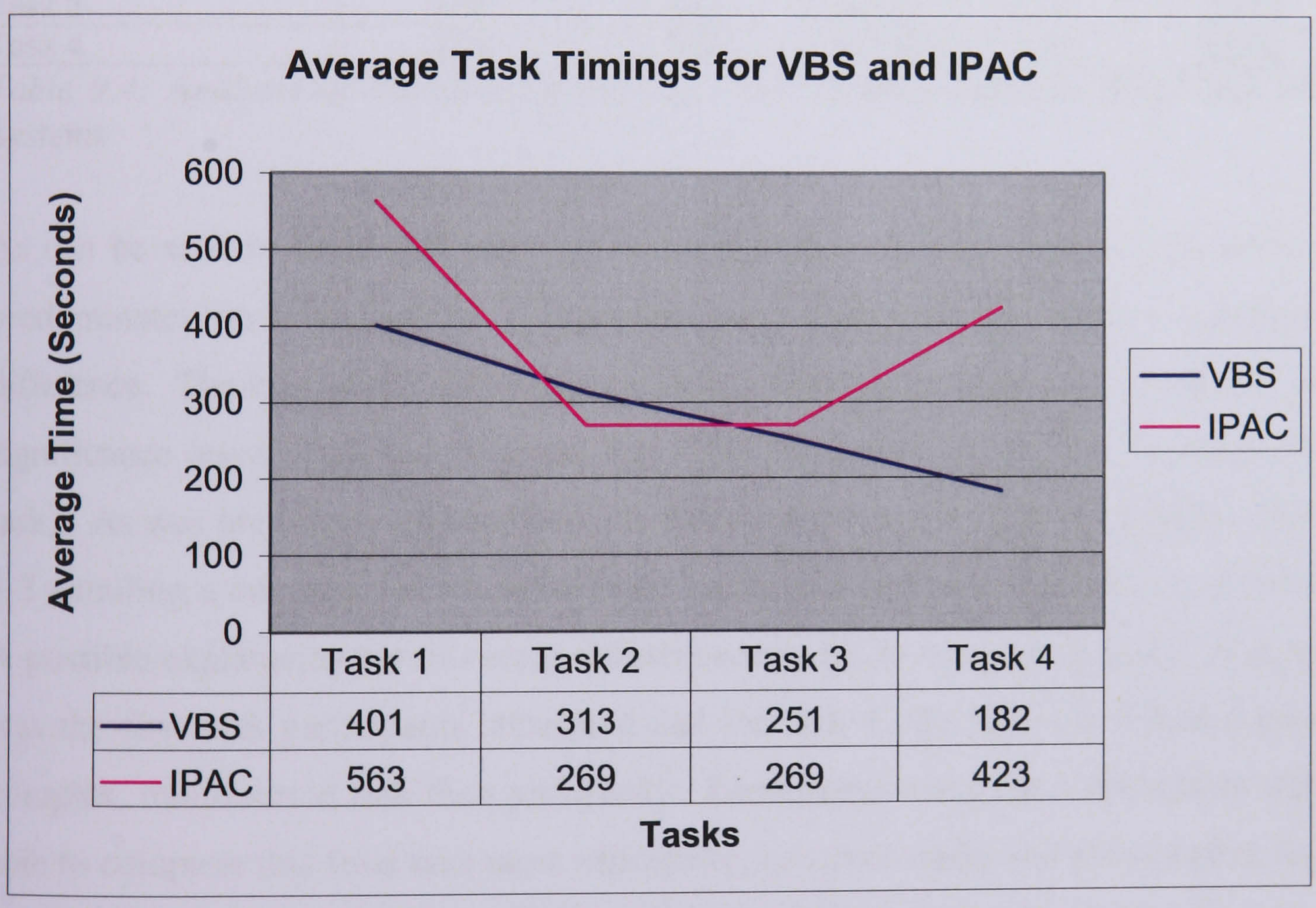


Figure 9.3: Graph Showing Timing Improvement over Task Set

It can be seen in Figure 9.3 that participants using the VBS system improved (reduced) their level of effort in terms of time taken as the tasks progressed; whilst users of the IPAC system, although slightly more efficient on Task 2, did not display this developmental pattern. This suggests that as users become more familiar with the View-based system, they become more efficient in the completion of complex

multifaceted subject searching tasks; a profile not evident in the specification-led IPAC interface. This relates to the study by Yee at al (2003) who found that with experience participants became "more comfortable" with the facet-based interface, although in contrast to this current investigation they actually discovered a significant difference in task timings in favour of the baseline, non-faceted interface (although they explain that this could be due longer system processing times). It can also be seen from Figure 9.2 that the apparent source of this significant difference in average task timings was tasks 1 and 4. This task timing data was submitted to a one-way analysis of variance as shown in Table 9.4 below:

Measure	IPAC Mean Time (Seconds)	VBS Mean Time (Seconds)	F	df	p
Task 1	563	401	3.761	1,31	0.062
Task 2	269	314	0.016	1,32	0.899
Task 3	269	251	0.114	1,32	0.738
Task 4	423	182	10.003	1,32	0.003

Table 9.4: Analysis of Variance of Average Task Timings between IPAC and VBS Systems

As can be seen in Table 9.4, although there is a clear difference in all task timings predominately in favour of the VBS system, only Task 4 demonstrated a significant difference. There is a clear difference in Task 1 timings but this was just outside the significance level (it should be noted that VBS Participant 1 failed to complete this task). As was previously exemplified, the nature of the 4 tasks was comparable; tasks 1-3 entailing a combination of two subject facets and task 4 a combination of three. A possible explanation for this clear difference may be, in the case of task 1, that this was the first task participants attempted and for task 4, the fact that it was a more complex, multifaceted task than previously. Participants using the VBS system were able to complete this final task more efficiently than previously and also quicker than the IPAC participants. This was perhaps facilitated by a greater level of experience, over and above the task complexity. Conversely, this greater task complexity may have hindered task completion for participants using the keyword-based IPAC. This longer timing is particularly noticeable in comparison to the prior tasks 2 and 3, in which IPAC participants displayed identical average task timings of 269 seconds. The above differences in task timings might be further investigated and clarified by future research that varies the task order with different participant groups. This could

entail repeating the experiment and dividing the VBS and IPAC groups into subgroups with different task orders to help clarify whether the actual task influenced the result or its position in the task sequence.

Following conclusion of the quantitative analysis, data were collated for qualitative investigation. There were two forms of qualitative data available in the data set: one set of data came from question 11 of the usability questionnaire, which asked people for their views on the experience of the tasks, paying particular attention to the interface style. These responses are recorded for reference purposes in Appendix 4. The second data set were obtained from session logs, which recorded key activities and search results obtained during performance of each task. As in the previous case, these data were largely open-ended and required qualitative investigation.

In relation to the question asking for comments about the library system, it was evident that VBS users encountered initial difficulties with the search style but all recognised the potential of the application. For example, participant 10 said:

> *I think once you familiarise yourself with the system, you feel comfortable, there is some need for tutorial, though a short one, for potential searches to use the system to its potential.*

Similar thoughts were voiced by participants 5, 13 and 15, respectively:

> *Once we understood how to use it, it was quite easy to use.*
>
> *...With a bit more practice, I would really enjoy using it and find it very efficient. Enjoyed it! Thank you!*
>
> *The above evaluation is in the light of having needed a practice go, in which I got a bit lost. Once I had had a go and learned where I'd gone wrong, I was confident in using the system and found it easy to find my way around.*

The above statement reiterates the previous task timing finding in that users demonstrated a reduction in effort levels (time taken) as they gained experience, progressing through the tasks. Again, this participant directly expresses a key finding of similar research by Yee et al (2003), in that "as participants continued to use the interface, they became more comfortable with it."

Interestingly, more experienced participants encountered little initial difficulty:

I used to be a librarian so am used to the Dewey system and have some understanding of the structures behind the system so I adapted fairly well to this new system. (Participant 7)

I am reckoning that this system is built on an SQL queried database. As I am competent in SQL, it was easy to translate these skills onto this system. Perhaps this helped when calculating what the headings and subheadings I needed were. This system made it easier to find information than the current system but some headings were too specific... (Participant 14)

Others actively 'discovered' information:

> *This system would be extremely helpful for people who had no idea what particular books they were searching for. I have found a few titles which I had not heard of which would be particularly useful for me.* (Participant 8)

There were weaknesses inherent in the VBS pilot system, particularly in relation interface elements. For example, those participants who preferred to instigate their search by entering terms into the search index (as opposed to selection by category), often failed to realise that, before hitting the Return key, they should select a term from the drop-down list box.

> *...When you do a search more explicit instruction would be better, i.e. type in a word NOW HIGHLIGHT ITEM IN POP UP LIST* (Participant 6)

Similarly, one of the perceived weaknesses of the VBS system was the lack of a tutorial and/or online help:

> *Would need a good set of instructions with step-by-step guidance and several practice examples.* (Participant 16)

> *The system gives you no help – if you get stuck it's up to you to backtrack...* (Participant 12)

A few participants experienced problems with the Dewey terminology or structure, when attempting to narrow their search:

> *I think I need more categories to narrow my search* (Participant 1)
>
> *...Categories for refinement sometimes not helpful* (Participant 4)

As was discussed in Chapter 5, the Dewey Summaries (top three hierarchical) used in the VBS system had been re-cast by OCLC for the purposes of online searching. The

above comments from users perhaps suggest that if the Dewey structures are to be used for online searching, further attention needs to be given to providing meaningful captions, particularly at lower hierarchical levels.

In relation to the alternative IPAC system, participants called for more system information about library items - "*…a short/small description of books*" (Participants 18 and 34), "*..definitely add abstracts to the different works listed*" (Participant 27) – whilst other participants (20 and 32) called for spell-check algorithms. Interestingly, a number of participants were seeking the ability to search laterally, linking to related items, an ability inherent in a grouping by classification:

> *It would be nice if it offered "similar" results or "guessed" what you were looking for – suggestions based on topic/subject ("see also"). A grouping by subject?* (Participant 19)
>
> *…Would be helpful if the system could make more lateral suggestions, particularly when searching unfamiliar subject areas* (Participant 23)
>
> *…Try building a subject index with a root system* (Participant 27)

Thus, in general, the qualitative analysis of the attitudinal data reinforced the finding that the VBS system offered enhanced benefits over the IPAC system for complex multifaceted subject searching tasks.

This finding was also reflected in analysis of system search logs, with IPAC participants seemingly experiencing many of the well-documented subject search problems, highlighted in Chapter 3, such as a difficulty in translating the search task into a conducive keyword search, entering too broad a search term and retrieving a large result set, or too narrow a term or a misspelt word to retrieve nothing at all. Table 9.5 shows the diverse range of search terms used by participants during Task 1, coupled with the search index used, number of participants using these terms and results retrieved (similar tables for the remaining three tasks can be found in Appendix 5).

Search Terms	Index	Number of Participants Using Terms	Number of hits
higher education west Yorkshire	General Keyword	9	1
Higher education institutions	General Keyword	4	135
higher education institutions west Yorkshire	Subject Keyword	4	0
higher education institutions yorkshire	General Keyword	3	1
west yorkshire education	General Keyword	8	22
west yorkshire universities	General Keyword	2	0
University west Yorkshire	General Keyword	2	53
huddersfield university	General Keyword	1	2727
huddersfield university institution	General Keyword	1	2
leeds university	General Keyword	1	204
Higher education	General Keyword	5	2688
universities,coleges, yorks	General Keyword	1	0
universities,yorks	General Keyword	1	0
universities	General Keyword	2	1043
Higher education yorkshire	Subject Keyword	1	1
Further Education Funding Council , Inspectorate	Author Keyword	1	237
Leeds University	General Keyword	1	204
higher education institutions in west yorkshire	General Keyword	6	0
University of Huddersfield, planning and resources group	General Keyword	1	78
Leeds University	Title Keyword	1	27
West Yorkshire universities	Title Keyword	1	6
Higher education establishments West Yorkshire	Title Keyword and Subject Keyword	1	0
Higher education reviews West Yorkshire	Title Keyword and Subject Keyword	1	0
West Yorkshire	Subject Keyword	1	321
Yorkshire - Local Authority Education Service	Subject Keyword	1	1
Education West Yorkshire	Title Keyword and Subject Keyword	1	7
"university + west Yorkshire"	General Keyword	1	0
Uk universities	Subject Keyword	1	0
Universities	Subject Keyword	1	594
Higher education institutions - Organisation and administration	Subject Keyword	1	45
Higher education institutions Yorkshire	Subject Keyword	1	0
Bradford university	Subject Keyword	1	4
Leeds University	Subject Keyword	1	4
Higher education institutions	Subject Keyword	1	85
HEI West Yorkshire	Subject Keyword	1	0
Higher Education Institutions WY	General Keyword	1	0
Higher Education Institutions West	General Keyword	1	0
Yorkshire and Humberside Association for Further and Higher Education	Author Keyword	1	5
Humberside	General Keyword	1	80
West Yorkshire	Subject Keyword	2	397
Higher Education West Yorkshire	Journal Title Keyword	1	0
Higher Education West Yorkshire	University Papers Keyword	1	7
Higher education west Yorkshire	Title Alphabetical	1	0

Higher education west Yorkshire	Title Keyword	1	8
Review west Yorkshire	Subject Keyword	1	0
Review Higher Education	Subject Keyword	1	2
Review Higher Education	Journal Title Keyword	1	1
Higher education review	Title Keyword	1	72
literature review about higher education institutions in west yorkshire	General keyword	1	0
Literature review	General keyword	2	160
literature review in west yorkshire uk	General keyword	1	0
West Yorkshire	General keyword	2	418
higher edication insitutions in west Yorkshire	General keyword	1	0
higher edication institutions	General keyword	1	0
Literature review higher education institutions	General keyword	1	0
higher education & west Yorkshire	General keyword	1	0
lierature review about higher educations	General keyword	1	0
lierature review in west Yorkshire	General keyword	1	0
lierature review about higher education	General keyword	1	0
lierature review	General keyword	1	0
Literature review about higher education	General keyword	1	0
literature review about higher education in west Yorkshire	General keyword	1	0
Higher education institution west Yorkshire	General keyword	2	1
Higher education	General keyword	1	12
Literature review of higher education in west Yorkshire	General keyword	1	0
Literature review of higher education west Yorkshire	General keyword	1	0
higher education in west Yorkshire	General keyword	1	2
Higher education in west Yorkshire	General keyword	1	1
Education in west Yorkshire	General keyword	1	10
highereducation in west Yorkshire	General keyword	1	0
educational institutes in west Yorkshire	General keyword	1	0
west yorkshire institutions	General keyword	1	1
west yorkshire svhools	General keyword	1	0
west yorkshire schools	General keyword	1	8
higher education schools in west Yorkshire	General keyword	1	0
education schools in west Yorkshire	General keyword	1	3
Higher education institution	General keyword	1	9
West Yorkshire education institute	General keyword	1	3
yorkshire he institute	General keyword	1	0
West Yorkshire institute	General keyword	1	23

Table 9.5: IPAC Search Terms, Index, Usage and Results for Task 1

The above Table 9.5 gives an indication of the diverse range of search terms used by the 17 IPAC participants, with a General Keyword search for "higher education west Yorkshire" being the most frequently used search terms, by 9 participants, yet retrieving just 1 library item. As can be seen from Table 9.6 below, for Task 1. 35 (47%) of the 80 different search terms entered by participants retrieved no results and that of a total of 119 searches by the different participants, 44 (37%) resulted in an

empty result set. As shown below, similar results were obtained for Tasks 2 and 4, yet with fewer total search terms, total searches and zero 'hits' for Task 3, possibly due to the strengths of the library collection and a better matching of search terms in this subject area. Across all four tasks, 38% of search terms and 37% of searches by participants returned zero results.

Task	Total Search Terms	Zero 'Hit' Search Terms (% of Total Terms)	Total Searches	Participant Searches with Zero 'Hits' (% of Total Searches)
1	80	35 (47%)	119	44 (37%)
2	64	35 (55%)	92	42 (46%)
3	48	8 (17%)	62	8 (13%)
4	97	32 (33%)	95	44 (46%)
TOTAL	**289**	**110 (38%)**	**368**	**137 (37%)**

Table 9.6: Number of IPAC Search Terms and Participant Searches with Zero 'Hits'

This concurs with OPAC studies by Markey (1989), cited in Chapter 3, which found that one third to one half of subject search terms returned no results. Moreover, it is also perhaps worth noting that 27 (61%) out of these 44 zero result searches for Task 1 were conducted using the broadest search index (General Keyword) with more scope to match than Markey's subject search index analysis, as it retrieves results from all indexes simultaneously, yet participants still experienced a search term failure rate (zero results) of between 33% (one third) to 55% (over half) for 3 out of the 4 tasks. This also relates to the study by Yee at al (2003) which found that participants using the keyword interface experienced zero results on 82 occasions, compared to 26 for the facet-based interface. As will be demonstrated by the VBS search profile, a concurrent result was found in the present study, with only 2 participants experiencing zero (empty) result sets for Task 1 (compared to 44 instances for IPAC participants) and 7 participants across the entire task set.

In addition to a high proportion of zero 'hits', the IPAC participants also appeared to realise that the type of index selected affected the search results, evident in the qualitative comments, here relating to Task 2:

> *Inconsistency – the search brought up topics title "English poets – 19th century" however if keying into one search file the same criteria, the computer could not find any search results* (Participant 25)

Across the four tasks, but particularly Task 1, IPAC participants were extremely individualistic in their choice of search terms and search style, as a group creating a more diverse search profile than their VBS counterparts and experiencing a greater number of zero 'hits'.

The VBS participants displayed a stronger, more cohesive search profile.

Task	Term Selection	Number of Participants Selecting Term
1	West Yorkshire England Northeastern England Kirklees Higher Education	12 1 1 2 16
Number of Zero Hits		2
2	Arts, Recreation, Sports Music Biography (Reference Facet) Persons Facet	1 14 5 9
Number of Zero Hits		3
3	English Poetry 18th Century Persons Facet Type of Library Item Facet	17 16 17 1 1
Number of Zero Hits		2
4	English Literature Fiction/Novels Women 19th Century Modern Period 1500-1899	14 15 10 17 16 1
Number of Zero Hits		0

Table 9.7: VBS Search Term Selection by number of Participants

Although not directly comparable to the previous Table 9.6, due to the difference in interface and search interaction, Table 9.7 shows that participants were able to 'find' the appropriate search terms and combine facets to complete the four tasks, retrieving fewer zero 'hits' (7 instances) than the IPAC participants.

In Task 1, all but one participant browsed or searched the top level 'Economics, Education, Society' Main Class to find Higher Education, and these 16 participants (1

participant failed to complete the task) then succeeded in combining this with the Place facet; however, 2 participants remained at hierarchical levels above West Yorkshire (England, Northeastern England) and 2 browsed beyond, selecting Kirklees. Only 2 participants experienced zero search hits.

Similarly, in Task 2, the majority of participants (14, 82 %) succeeded in browsing or searching the Main Class Arts, Recreation and Sports to discover Music. Interestingly, the participants then pursued different strategies; only 5 selected the expected Reference facet to find biography, whilst 9 combined Music with the Persons facet to retrieve similar library items. As reflected in the previously discussed qualitative comments and quicker task timings, as participants used the VBS system they became more comfortable and efficient. This was evident in the transaction logs for Tasks 3 and 4.

In Task 3, all 17 participants selected English from the Literature Main Class and 16 then discovered poetry; all participants combined this with the 18th century within the Time facet. The number of participants experiencing zero hits reduced from 3 to 2.

The final task was arguably the most complex, having four subject components (English, fiction/novels, female/women, 19th century). As shown in Table 9.7, all 17 participants were able to narrow their search by the Person facet (women) and the Time Facet (19th century), although 1 participant browsed at the level above (Modern Period 1500-1899). However, participants did manage the subject discipline component in slightly different ways. The majority selected the Main Class Literature, followed by English but only 10 of these 14 participants then restricted their search to Fiction. As such, other participants tended to browse library item titles by subject discipline at a broader hierarchical level than the anticipated English fiction (e.g. English Literature or simply Literature), yet with a restriction by the Person and Time facets, which created a manageable result set. Thus, participants perhaps felt that they had narrowed their search sufficiently. Encouragingly, there were no instances of zero hits for this final Task 4.

9.5 Conclusion

The intention of this trial was to provide preliminary validation of the four derived facets and View-based Searching. The data analysis showed quite conclusively that in the context of this experiment a VBS system provides a viable alternative to a traditional form fill-in OPAC and in certain situations may even provide enhanced benefits. For example, the data analysis revealed that as people became more familiar with facet-based search strategies their task accomplishment times decreased. Furthermore, participants using the View-based system demonstrated a significantly lower average task time than IPAC system participants, suggesting a more efficient completion of multifaceted subject search tasks when using a View-based approach, even more so as experience is gained. This finding contributes to the work of Yee et al (2003), who found that with experience users displayed actions that suggest+ed they were "more comfortable" with a facet-based interface; however, task timings conversely displayed a significantly greater effort level, in comparison to the baseline keyword interface; this latter finding therefore contrasting with that of the current study. Participants using the VBS interface also displayed a more cohesive search profile and retrieved fewer zero (empty) result sets than the IPAC system, corresponding with the findings of Yee et al (2003). In addition, the experiment also demonstrated that people were positively disposed to facet-based searching and attitudes strongly reflected this. These findings concur with previous research (Yee et al, 2003), which concluded that a facet-based search approach was both successful and strongly preferred by participants.

The current findings could be further explored and validated in future research. Future research could adopt a within-subjects design to enable a more direct system comparison and a focus upon user preference. Another fruitful research design would be a repeated measure, particularly since the current findings suggested that system experience was an influential factor. Such a design would enable system testing with novice users and then revisiting this set of users, following accumulated experience. Testing the VBS system with the addition of a help file and tutorial might also prove interesting. Another variable worth exploring may be task choice. In the current research, the chosen tasks were all multi-faceted subject searches; further research

could explore simple, single facet subject tasks. This would investigate and validate Yee et al's (2003) finding of 50% of participants preferring the baseline keyword for single facet tasks, compared to 88% preferring the facet-based interface for their typical search type and 91% preferring it overall. Investigation into the relationship between search type (single/multifaceted), interface style and perceived search success may therefore be an area worth exploring. The current task timing results could also be further explored and clarified in a more controlled experiment, investigating the influence of varying task orders on task accomplishment time. More detailed analysis and capture (e.g. video) of the user-interaction, incorporating think-aloud user protocols, might also be a rich source of data, revealing user search strategies, navigation paths, problems and perceived/actual success.

CHAPTER 10

CONCLUSION

The research has argued that the intellectual effort of traditional library classification, the oldest knowledge organisation tool in existence, is currently under-utilised. Despite once used for the dual purpose of physical organisation of library collections and information retrieval in printed library catalogues, classification is now relegated to a shelf-location device. An historical legacy of pioneering hierarchical enumerative classifications of the nineteenth century, primarily the Dewey Decimal Classification Scheme, was embedded and later entrenched, despite societal changes in knowledge production seeking more specific and malleable faceted classifications. Although appending some synthetic devices, there became a growing disparity between the predominant classificatory paradigm and the universe of knowledge it sought to represent. The prevalence of these traditional hierarchical and enumerative classifications was then underlined by drives for standardisation, emanating from the development of the library catalogue in the United States. The classified catalogue was abandoned, although preferred in Europe, in favour of the alphabetic dictionary catalogue, which became the standard and model for the Online Public Access Catalogue. From this relegation of classification and the influence of batch-mode computer systems, not intended for a novice end-user, emerged an OPAC conceived, and still primarily adhering, to finding what is known not promoting exploration and discovery. Subject searching problems were well documented from the outset, with users experiencing difficulties in matching their search terms to the unseen vocabulary of the system, retrieving far too much or nothing at all. The value of a classificatory approach is being recognised by the Internet community for the organisation and retrieval of digital resources and in recent developments to create metadata for the Semantic Web. Although the initial drive was for vocabulary standardisation and interoperability, facet-based interfaces are beginning to emerge. However, commercial OPACs are still adopting a model of specification led, form fill-in interfaces, which are incongruous with the exploratory nature of subject searching and do

little to encourage multidimensional browsing. Problems still remain. There is a wealth of intellectual effort encoded within classificatory notations, which could be exploited at the interface for providing such enhanced subject access, yet this remains undiscovered. The Internet community is beginning to realise and exploit the value of classification, and increasingly faceted classification, for providing browsable, multidimensional discovery of digital resources. Yet, paradoxically, library OPACs maintain a restrictive 'finding' model, much to the detriment of their end users.

The present research aimed to address problems of subject access in library OPACs by manipulating a rigid, yet most widely used traditionally enumerative and hierarchical scheme: the Dewey Decimal Classification. In the course of the investigation, Dewey's hidden facets - for display at a system interface for multidimensional searching and browsing – were revealed. Prior to this work, both physical and the limited online applications of Dewey had displayed a library collection by its primary facet, subject discipline, with the remaining facets left undiscovered and scattered across the classificatory notation and library collection. Replication of the linear restrictions prescribed by the physical domain is clearly unnecessary in the domain of online information retrieval and this research freed classification from the 'shackles' imposed by physical arrangement, and its need for a prescribed citation order, by exploiting the multidimensional potential of classification in an online domain. In particular, it has laid the foundation for alternative classificatory arrangements at the OPAC user interface, readily tailored to individual needs.

To achieve this, a novel procedural method was devised, implemented and evaluated to deconstruct Dewey into four fundamental facets, which were populated with a library collection for display at a View-based Searching OPAC interface. The extent to which the derived facets assisted users' subject searching perceptions and activities was investigated in a small-scale usability study. The results provide strong evidence to support the idea that the derivation and population of such facets is achievable and that users are receptive to facet-based searching at the OPAC interface. In particular, the

facet-based system performed equally as well as a current form fill-in interface and, in certain cases, provided enhanced benefits.

The following sections identify contributions to knowledge and provide suggestions for future research. The thesis concludes with some closing thoughts.

10.1 Contributions

The derivation of appropriate, generic subject facets presents a considerable challenge to any research aimed at testing the idea of facet-based searching, particularly since the classification scheme under investigation has its origins in nineteenth century enumeration, with its facets embedded within a pre-coordinated notation. In order to guide the requisite deconstruction process of the most widely used bibliographic classification scheme in existence, the Dewey Decimal Classification Scheme, it was necessary to devise, successfully implement and evaluate a novel procedural method for deriving and populating fundamental facets. Now specified, the general method can be applied to any library collection that has been classified by Dewey, deriving fundamental facets specifically chosen to be generic to a library user from any subject area. Even if another library requested an alternative choice of facets, the same general procedure would apply, only with adjustments to the DDC criteria, since a different set of concepts would be represented by different notational components within Dewey. The research also encountered, highlighted and addressed problematic inconsistencies within Dewey, primarily related to the lack of a uniform notation. Thus, whilst developing the method, as practical problems were experienced, solutions were developed that incorporated criteria into the procedural method to help resolve such difficulties (for example, see exclusion heuristics, Chapter 6). The method, although underpinned by numerous SQL queries, ultimately allows a facet to be populated with library items using a single union query.

The research addressed constraints of physical library arrangements - the need for a citation order and resultant disciplinary scatter - by manipulating classification schemes

in the online domain. The derived facets collate library items according to the generic dimensions of time, place, reference type and person type, which would otherwise be embedded within a classificatory notation and scattered across a library collection. Such fundamental facets are implicit in multiple library items (have a strong literary warrant) and are desirable subject access points for many library users, regardless of subject discipline. Providing alternative arrangements by such embedded components has served to move the application of both physical and online bibliographic classification schemes from a predominantly linear to multidimensional platform, exploiting the faceted potential of a traditionally enumerative and hierarchical scheme such as Dewey.

When applied at a View-based searching interface, the derived facets provide additional subject access points and arrangements for a library user, which has proven advantageous to participants of this current research. The usability study results indicated that participants were positively disposed to subject searching with the derived facets. This contributes to the previously described (Chapters 3 and 9) research by Yee et al (2003), which investigated and demonstrated the potential of facet-based searching for image retrieval. Although Yee at al (2003) found that with experience participants became "more comfortable" with the interface, they actually discovered a significant difference in task timings in favour of the baseline, non-faceted interface (although they explain that this could be due longer system processing times). Conversely, this current research demonstrates a significant difference in favour of the faceted (View-based) interface in terms of overall average task timings and also an improvement (reduction) in participants' task completion times over the task set, as experience was gained. Such a positive learning effect was not observable for participants who completed tasks using the current form fill-in OPAC. Therefore contributing to Yee et al's (2003) research into image retrieval, the current research validates the potential of facet-based interfaces for subject searching within the domain of the library Online Public Access Catalogue.

The research also has implications for those working towards the vision of a Semantic Web. Core to this vision is the explicit representation of meaning, usually with the aid of ontologies (formal specifications of concepts/vocabularies). This vision will never be

manifested without effective tools to assist in the generation of these 'explicit representations of meaning' from existing structured information sources. The current research has demonstrated how semantic web researchers might unlock the Dewey system and extract meaning from the codes. Utilising vocabularies such as SKOS Core, this may make the automated generation of semantic mark-up for library items a step closer

10.2 Future Research

The research described to this point has provided an insight into the problems associated with the design of user-centred online retrieval systems for the library domain and has discussed, at length, strategies to facilitate subject searching by browsing. Moreover, a range of tools and techniques have been outlined which might usefully be employed to exploit and combine the intellectual effort of librarians in assigning classificatory notations with the advantages that facet-based searching offers to the less skilled user. However, it would be naïve to imagine that in its present form the method is sufficiently robust and that the case for the usability of facet-based searching has been proven. In order to achieve this, further work will need to be undertaken.

Further Testing of the Procedural Method

The method for facet derivation could be validated against other library collections that have been classified by the DDC. This would enable the identification of any additional inconsistencies within Dewey, pertaining on facet derivation and population, helping tailor the SQL query criteria accordingly.

Derivation of Additional Facets

The Standard Subdivisions and other DDC Tables could be investigated further to identify other candidate facets, which might be usefully incorporated.

Investigation of Dewey 'Bookmarks'

It was mentioned in the previous Chapter 9 that, although not the focus of the current research, specific subject views could be created according to the needs of a user population, e.g. by university course – Textiles, Hospitality Management etc. Such 'bookmark' views entail the display of multi-disciplinary subject areas, which may otherwise be scattered across the Schedules or buried at a deeper hierarchical level, collocating and promoting these within a single view. This could provide a less labour intensive supplement to Dewey deconstruction, providing a tailored user profile view that is activated on system entry by individual log-on details.

Supplementary Use of Library of Congress Subject Headings

Library of Congress Subject Headings are a controlled vocabulary of index terms assigned to library item bibliographic records, representing an internationally recognised standard. The scheme does not have a browsable classificatory structure, equivalent to the DDC, but it could be utilised in conjunction with this as an additional tool for facet population. The LCSH scheme is characteristically specific, whereby terms are chosen which are co-extensive with the subject not broader, unlike some instances of DDC application, e.g. the Rule of Three, whereby a multidisciplinary library item (three or more interdisciplinary subject areas) is classified, and specificity compromised, according to an inclusive yet broader number (see Tinker et al, 1999). LCSH is also synthetic in application, in that components such as the subdivisions are combined with main headings by the cataloguer, according to a set of rules. Main headings and subdivisions are coded accordingly within a bibliographic MARC record.

The first component of an LCSH index string, the main heading, can refer to the topic (tag 650), form of work (tag 650), geographic place (tag 651) or a personal, corporate or meeting name (tags 600, 610, 611), each of which is denoted as such in the MARC coding, as the figures in brackets indicate (Library of Congress, 2005a). This MARC coding may potentially prove very useful for facet population since it delimits individual subject code components and identifies their particular type. For example, a MARC code

could express a type of work (e.g. encyclopaedia) or a geographic area, which in this case would be as a primary topic, similar to an entry in the Dewey Schedules.

Main headings can then be qualified to increase their specificity by assigning subdivisions. There are 4 types of subdivision each of which are signified in the MARC record by way of subfield codes:

Topical($x)
Form($v)
Chronological($y)
Geographic ($z)

For a book entitled *Indigenous art: art gallery of Western Australia*, the above subfields are encoded within the 650 MARC field, as follows (Library of Congress, 2005b):

650 _0 |**a** Art, Aboriginal Australian |**y** 20th century |**v** Catalogs.
650 _0 |**a** Art |**z** Australia |**z** Perth (W.A.) |**v** Catalogs.

As the above demonstrates these subdivisions show a good correspondence with the facets established in the current research, primarily geographic, type of material (form) and chronological (dates and historical periods). These LCSH subdivisions are therefore another source (in addition to the Dewey Schedules, Standard Subdivisions and item titles) for identifying library items to populate the facets. Unlike Dewey where a transition between subject components is often not explicit, in LCSH these subdivisions should be readily identifiable using the MARC coding in the bibliographic record. This MARC coding is sadly lacking for the DDC, despite being proposed by Wajenberg (1983).

In addition to this MARC coding, WebDewey (an online tool for classifiers) (OCLC, 2004a) and the DDC homepage (OCLC, 2004b) provide links from LCSH headings to appropriate Dewey codes. These DDC-LCSH mappings may provide a means for assigning multiple Dewey codes to library items in the electronic domain, unconstrained by the remit of a single physical shelf location.

Comparison with the Universal Decimal Classification Scheme

It would be interesting to investigate a similar facet derivation and population with a more malleable scheme such as the UDC. As discussed in Chapter 2, although now not widely used, this scheme clearly delineates facet type and boundaries.

Further evaluation of Facet-based searching

The evaluation undertaken in this research has comprised a small-scale usability study. It was recommended in the previous Chapter 9 that this be validated with additional evaluation, as follows:

- Alternative research designs (within-subjects, repeated measures)
- Testing with the addition of a tutorial and help file
- Comparison of current multi-faceted tasks with simple, single facet tasks
- More detailed capture and analysis of user-system interaction

Dewey Deconstruction and the Semantic Web

The current research has analysed and derived facets of Dewey to improve subject access to physical library collections in library OPACs. However, this deconstruction of Dewey might also be applied and extended to resources on the Web, providing a way forward for an explicit representation of meaning for the DDC and its associated items. Drawing upon vocabularies such as SKOS Core, this could facilitate the automatic semantic mark-up of library items, incorporating numerous pre-existing and pre-structured resource collections into a more fully realised vision of a Semantic Web.

10.3 Concluding Remarks

This research has demonstrated how the readily available notation of the primarily enumerative, hierarchical Dewey Decimal Classification scheme can be manipulated to achieve a faceted, multidimensional search platform for enhanced subject retrieval in library OPACs. Owing to its rigid hierarchical structure, pre-coordinated and sometimes inconsistent notation and lack of distinct facet indicators, the DDC would perhaps not be the first, or easiest, choice for the implementation of such faceted retrieval. Yet it constitutes a *de facto* international standard, a scheme which, like classification *per se*, is

now under-utilised for subject retrieval in our commercial library OPACs, relegated to a role of shelf location.

If we are to address the continuing problem of subject access, we perhaps need to reconsider the wealth of intellectual effort in our bibliographic records and how this can be best stored to enable effective subject browsing and discovery at the OPAC interface. A classificatory notation is pre-coordinated for the purposes of shelf arrangement, Rananathan's multidimensional universe of knowledge spread "...along one line" (Ranganthan, 1965, p.198), yet this notation, as Wajenberg (1983) recommended over twenty years ago, could be stored in its component parts. If we are to exploit our dormant classification schemes at the interface, then deconstruction is an act of reverse engineering if these schemes are stored as a pre-coordinated string. OCLC, the publishers of Dewey, are already investigating usage of such post-coordinated Dewey notation storage by identifying the component parts of synthesised numbers in a 'hidden number' tag (Beall, 2003). However, this facility is currently only available in their Editorial Support System, not in bibliographic MARC records for the end user. Such classificatory component storage and, primarily, its subsequent exploitation would be worthy of further investigation and extension to the OPAC end-user, helping utilise the richness of classificatory notations and bringing more innovative interfaces to our library systems.

If library OPACs persist in maintaining the legacy of the specification-based dictionary catalogue, whilst the Internet community discover and increasingly exploit the potential of classification, particularly faceted classification, for searching and browsing then the problems of subject access in library OPACs may remain, with end-users becoming gradually more disheartened as to the growing disparity between these two online 'worlds'. The library community should be encouraged to re-discover and investigate how to exploit the potential of their oldest knowledge organisation tool. Only then can classification be taken beyond the linear constraints of physical shelf arrangement towards a multidimensional application at the online interface, promoting subject browsing and discovery in our library OPACs.

References

Aduna (2005) *Welcome to the DOPE browser!* [Online] Available at: < http://aduna-software.com/index.html> [Accessed 30 June 2006].

Allen, B. L. (1991) Cognitive research in information science: implications for design. *Annual Review of Information Science and Technology* 26. Washington: American Society for Info Science, pp. 2-37.

Allen, R. B. (1996) *Retrieval from facet spaces.* [Online] Available at: <http://www.raven.umd.edu/~rba/PAPERS/FACETS/facets.html> [Accessed 21 February 2005].

Artifact (2004) *Aftifact: the arts and creative industries HUB. Cataloguing guidelines.* [Online] Available at: <http://www.artifact.ac.uk/documents/artifact-cataloguing-guidelines.htm#2> [Accessed 18 January 2005].

Arts and Humanities Data Service (2005) *Arts and Humanities Data Service* [Online] Available at: <http://www.ahds.ac.uk> [Accessed 4 April 2005].

Baeza-Yates, R. and Ribeiro-Neto, B. (1999) *Modern information retrieval.* Harlow: ACM Press.

Baker, S. L. and Lancaster, F. W. (1991) *The measurement and evaluation of library services*, 2nd ed. Arlington, Virginia: Information Resources Press.

Bates, M. J. (1977) System meets user: problems in matching subject terms. *Information Processing and Management* 13, 205-214.

Bates, M. J. (1986) 'Subject access in online catalogs: a design model' *Journal of the American Society for Information Science* 37(6), 357-376.

Bates, M. J. (1989) 'The design of browsing and berrypicking techniques for the online search interface' *Online Review* 13(5), 407-424.

Bates, M. J. (1996) *Indexing and access for digital libraries and the Internet: human, database, and domain factors* [Online] Available at: <http://www.gseis.ucla.edu/faculty/bates/articles/indexdlib.html> [Accessed 21 February 2005].

Bawden, D. (1986) Information Systems and the stimulation of creativity. *Journal of Information Science* 12(5), 203-216.

Beall, J. (2003) *Hidden numbers.* [Personal Communication, 9 July, 2003].

Beghtol, C. (1995) 'Facets'as interdisciplinary undiscovered public knowledge: S.R. Ranganathan in India and L. Guttman in Israel' *Journal of Documentation* 51(3), 194-224.

Beghtol, C. (1998a) General classification systems: structural principles for multidisciplinary specification. *In:* W Mustafa el Hadi, J Maniez and A S Pollitt (eds.) *Structures and relations in knowledge organisation: proceedings of the fifth international ISKO conference 25-29 August, Lille France.* Wurzburg: Ergon Verlag, pp.89-96.

Beghtol, C. (1998b) 'Knowledge domains: multidisciplinary and bibliographic classification systems' *Knowledge Organisation* 25 (1/2), 1-12.

Belkin, N. J. (1981) Ineffable concepts in information retrieval. *In:* Sparck Jones. K (ed.) *Information retrieval experiment.* London: Butterworths, pp. 44-58.

Belkin, N. J., Oddy, R. and Brooks, H. (1982) ASK for information retrieval: part I. background and theory *In:* K Sparck-Jones & P Willet (eds.) *Readings in information retrieval.* San Francisco, CA: Morgan Kaufmann, 1997, pp.299-304.

Belkin, N. J. (1990) 'The cognitive viewpoint in information science'. *Journal of Information Science* 16, 11-15.

Belkin, N. J., Marchetti, P.G. and Cool, C. (1993) 'BRAQUE: design if an interface to support user interaction in information retrieval' *Information Processing and Management* 29(3), 325-344.

Berners-Lee, T., Hendler, J. and Lassila, O. (2001) The Semantic Web. *Scientific American* May 17th 2001. [Online] Available at: http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2 [Accessed 22 March 2005].

Bibliographic Data Services Limited (no date) *Bibliographic Data Services Limited* [Online] Available at: < http://www.bibliographicdata.com/> [Accessed 21 February 2005].

BIOME (2004) *Browse BIOME.* [Online] Available at: http://biome.ac.uk/browse/ [Accessed 21 February 2005]

Bliss Classification Association (2002) *Progress of BC2 in 2001: editor's report to the AGM.* [Online] Available at: < http://www.sid.cam.ac.uk/bca/Report2001.htm> [Accessed 21 February 2005].

Bliss Classification Association (2003) *Progress of BC2 in 2003: editor's report to the AGM.* [Online] Available at: < http://www.sid.cam.ac.uk/bca/Report2003.htm> [Accessed 21 February 2005].

Bliss Classification Association (2004) *The Bliss Bibliographic Classification.* [Online] Available at: http://www.sid.cam.ac.uk/bca/bchist.htm [Accessed 21 February 2005].

Borgman, C. L. (1986) 'Why are online catalogs hard to use? Lessons learned from information-retrieval studies'. *Journal of the American Society for Information Science* 37(6), 387-400.

Borgman, C. L. (1996) 'Why are online catalogs still had to use?' *Journal of the American Society for Information Science* 47(7), 493-503.

Brooke, J. (1986) *SUS – A quick and dirty usability scale* [Online] Available at: < http://www.usability.serco.com/trump/documents/Suschapt.doc> [Accessed 4 April 2005].

Broughton, V. (1999) 'Notational expressivity; the case for and against the representation of internal subject structure in notational coding.' *Knowledge Organization* 26(3), 140-148.

Broughton, V. (2001a) *Theory and practice in faceted classification schemes. A two day course, 26th-27th June 2001, University College London* [Unpublished course notes].

Broughton, V. (2001b) 'Faceted classification as a basis for knowledge organization in a digital environment; the Bliss Bibliographic Classification as a model for vocabulary management and the creation of multi-dimensional knowledge structures.' *The New Review of Hypermedia and Multimedia* 7, 67-102.

Broughton, V. (2002) *Facet analytic theory as a basis for a knowledge organization tool in a subject portal. Paper presented at the Seventh International ISKO Conference: Challenges in knowledge representation and organization for the 21st century, integration of knowledge across boundaries, 10-13 July 2002, Granada, Spain.* [Online] Available at: <http://www.ucl.ac.uk/fatks/paper2.htm> [Accessed 23 August 2004].

Broughton, V. and Lane, H. (2004) The Bliss Bibliographic Classification in action: moving from a special to universal classification scheme via a digital platform. *Knowledge organization and the global information society: proceedings of the 8th international ISKO conference 13-16 July, London, UK.* [Online] Available at: <http://demo1.thingsprime.com:8080/prime/filesexporter?m1=12&rid=52&vnum=0> [Accessed 22 March 2005].

Buchanan, B. (1979) *Theory of library classification.* London: Bingley.

Campbell, D. J. (1977) A short biography of Henry Bliss (1870-1955). *In:* Mills, J and Broughton, V (eds.) *Bliss Bibliographic Classification, 2nd ed. Introduction and auxiliary schedules.* London: Butterworths, pp.1-9.

Carpenter, M. (1994) 'Catalogs and cataloging' *In:* Wiegand, W.A. & Davies, D G (eds.) *Encyclopedia of library history.* London: Garland, pp. 107-117.

Cleverdon, C. (1997) 'The time of Dorking' *In:* Gilchrist, A (ed.) *From classification to knowledge organization. Dorking revisited of 'past is prelude'. A collection of reprints to commemorate the 40 year span between the Dorking Conference (First International Study Conference on Classification research 1957) and the Sixth International Study Conference on Classification Research, London, UK)* The Hague: International Federation for Information and Documentation, p.vi.

Chan, L. M. (1986) 'Library of Congress Classification as an online retrieval tool: potential and limitations' *Information Technology and Libraries* 9, 181-192.

Chan, L. M., Comaromi, J.P. and Satija, M.P. (1996) *Dewey Decimal Classification: a practical guide.* 2nd ed. Albany, NY: Forest Press.

Chang, S. J. and R. E. Rice (1993) 'Browsing: a multidimensional framework` *In:* Williams, M E (ed.) *Annual Review of Information Science and Technology* 28. Medford, NJ: Learned Information.

Cimiano, P., Haase, P., Sure, Y., Völker, J. and Wang, Y. (2006) *Question answering on top of the BT digital library.* [Online] Available at: < http://www2006.org/programme/files/pdf/p175.pdf> [Accessed 30 June 2006].

Classification Research Group (1955) 'The need for a faceted classification as the basis of all methods of information retrieval' Gilchrist, A (ed.) *From classification to knowledge organization. Dorking revisited of 'past is prelude'. A collection of reprints to commemorate the 40 year span between the Dorking Conference (First International Study Conference on Classification research 1957) and the Sixth International Study Conference on Classification Research, London, UK)* The Hague: International Federation for Information and Documentation, 1997, pp. 1-9.

Cochrane, P. A. (1982) 'Classification as a user's tool in online public access catalogs' *Studien zur klassifikation* 11, 260-268.

Cochrane P. A. and Johnson, E. H. (1996) Visual Dewey: DDC in a hypertextual browser for the library user. *In:* Green, R (ed.) *Knowledge organization and change, proceedings of the fourth international ISKO conference, 15-18 July 1996, Washington DC, USA.* Frankfurt/Main: Indeks Verlag.

Comaromi, J. P. (1983) The foundations of the Dewey Decimal Classification: the first two editions. *In:* Stevenson, G and Kramer-Greene, J (eds.) *Melvil Dewey: the man and the classification.* Albany, New York: Forest Press, pp.135-147.

Cutter, C. A. (1904) *Rules for a dictionary catalog*, 4th ed. Washington: Government Printing Office.

Dahl, R. (1973) *Charlie and the chocolate factory.* Harmondsworth: Penguin.

Davies, J., Fensel, D and van Harmelen, F. (2003) 'Introduction'. *In:* Davies, J., Fensel, D. and van Harmelen, F. (eds.) *Towards the Semantic Web: ontology-driven knowledge management.* Chichester: John Wiley & Sons Ltd.

Davies, R. (1989) The creation of new knowledge by information retrieval and classification. *Journal of Documentation* 45(4), 273-301.

Dewey, M. (1996) *Dewey Decimal Classification and relative index. Vol. 1 introduction and tables*, 21st ed. Mitchell, J.S. et al (eds). Albany, New York: OCLC, Forest Press.

Dolby, R. G. A. (1979) 'Classification of the sciences. The nineteenth century tradition' *In:* Ellen, R F & Reason, D (eds.) *Classifications in their social context.* London: Academic Press, pp. 167-193.

EEVL (2004) *EEVL: collection development policy.* [Online] Available at: <http://www.eevl.ac.uk/cdp.htm> [Accessed 18 January 2005].

Efthimiadis, E. N. (1994) 'End-users' understanding of thesaural knowledge structures in interactive query expansion'. *Advances in Knowledge Organization* 4, 295-303.

eLera (2006) *E-learning research and assessment network.* [Online] Available at: <http://elera.net/eLera/Home> [Accessed 30 June 2006].

Ellis, D. (1989) 'A behavioural approach to information retrieval system design' *Journal of Documentation* 45(3), 171-212.

Endeca (2004) *About Endeca: The easiest way to find information.* [Online] Available at: <http://endeca.com/about_endeca/index.html> [Accessed 21 February 2005].

Facet Map (2003) *FacetMap: your home for faceted classification tools.* [Online] Available at: <http://facetmap.com/> [Accessed 21 February 2005].

Fluit, C., Sabou, M. and van Harmelem, F. (2006) Ontology-based information visualization: toward Semantic Web applications. *In:* Geroimenko, V. and Chen, C. (eds.) *Visualizing the Semantic Web: XML-based Internet and information visualization.* 2nd ed. London: Springer-Verlag Ltd, pp. 45-79.

Foskett, A. C. (1982) The subject approach to information, 4th ed. London: Clive Bingley

Foskett, A. C. (1996) The subject approach to information, 5th ed. London: Library Association Publishing.

Foskett, A. C. (2000) 'The future of faceted classification.' In: Marcella, R & Maltby, A (eds.) *The future of classification*. Aldershot: Gower.

Foskett, D. J. (1997) Dorking revisited. *In:* Gilchrist, A. D. (ed.) *From classification to knowledge organization. Dorking Revisited or 'past is prelude'. A collection of reprints to commemorate the 40 year span between the Dorking international conference (first international study conference on classification research, 1957) and the 6th international study conference on classification research (London, UK, 1997).* The Hague: FID, p.ix.

Fraser, M. and Edwards, A. (2001) *Humbul Humanities Hub: cataloguing guidelines.* [Online] Available at: < http://www.humbul.ac.uk/about/catalogue. html# introduction> [Accessed 21 February 2005].

Freeman, R. R. and Atherton, P. (1967) American Institute of Physics/UDC project AUDACIOUS: file organization and search strategy using the Universal Decimal Classification in mechanized reference retrieval systems. Excerpts from a paper presented at the F.I.D/I.F.I.P conference on mechnized information storage and retrieval, and dissemination in Rome, June 15 1967. *In:* Cochrane, P (ed.) *Redesign of catalogs and indexes for improved online subject access: selected papers of Pauline A. Cochrane.* Phoenix Arizona: Oryx Press, 1985, pp. 312-323.

Garshol, L. M. (2004) Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all. [Online] Available at: http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html

Geroimenko, V. and Chen, C. (2006) 'Preface' *In:* Geroimenko, V and Chen, C. (eds.) *Visualizing the Semantic Web: XML-based Internet and information visualization*, 2nd ed. London: Springer-Verlag Ltd. pp. v-vii.

Gillchrist, A. D. (1994) 'Classifications and thesauri' In: Vickery, B C (ed.) *Fifty years of information progress: a journal of documentation review.* London: Aslib, pp.85-118.

Goble, C. (2003) The Semantic Web: an evolution for a revolution. *Computer Networks* 42(5), 551-556.

Hancock, M. (1987) 'Subject searching behaviour at the library catalogue and the shelves: implications for online interactive catalogues' *Journal of Documentation* 43(4), 303-321.

Hancock-Beaulieu, M. (1990) 'Evaluating the impact of an online library catalog on subject searching behavior at the catalog and at the shelves.' *Journal of Documentation* 46(4), 318-338.

Hanson, T. (1998) 'The access catalogue: gateway to resources.' *Ariadne* 15 [Online] Available at: <http://www.ariadne.ac.uk/issue15/main/> [Accessed 22 March 2005]

Harter, S. P. and Hert, C. A. (1997) 'Evaluation of information retrieval systems: approaches, issues, and methods.' *In:* Williams, M E (ed.) *Annual Review of Information Science and Technology* 32. Medford, NJ: Information Today, pp.3-94.

Herman, I (2005) Questions (and answers) on the Semantic Web. *Keynote at Semantic Web Days conference, 6-7 October 2005, Munich, Germany.* [Online] Available at: http://www.semantic-webdays.net/proceedings/keynote_Herman_SemanticWebDays 2005.pdf [Accessed 30 June 2006].

Hert, C. A. (1996) *Information retrieval as situated action* [Online] http://www-slis.lib.indiana.edu/Research/chert-situated.html [Accessed 4 September 1998].

Hert, C. A. (1997) *Understanding information retrieval interactions: theoretical and practical implications.* 'London: Ablex Publishing Corporation

Hewins, E. T. (1990) 'Information need and use studies' *In:* Williams, M E (ed.) *Annual Review of Information Science and Technology* 25. Amsterdam: Elsevier, pp. 145-172.

Hildreth, C. (1995) *Browsing and exploring: a new paradigm for IR/OPAC system design.* [Online] Available at: <http://myweb.cwpost.liu.edu/childret/clr-five.html> [Accessed 22 March 2005].

Humbul (2005) *Humbul humanities hub.* [Online] Available at: < http://www.humbul. ac.uk/> [Accessed 22 March 2005].

Hunter, E. (2000) 'Do we still need classification?' *In:* Marcella, R & A Maltby (eds.) *The Future of Classification.* Aldershot: Gower, pp.1-17.

Hussain, A. and O'Brien, A. (1992) 'Recent trends in subject access to OPACs: an evaluation' *International Classification* 19(3), 140-145.

Hyvönena,E., Saarela, S. and Viljanen, K. (2004a) Application of ontology techniques to view-based semantic search and browsing. *In* C. Bussler, J. Davies, D. Fensel, R. Studer (eds.): *The Semantic Web: Research and Applications. Proceedings of the First European Semantic Web Symposium* (ESWS 2004), Springer-Verlag, LNCS 3053, 2004. [Online] Available at: <http://www.cs.helsinki.fi/u/eahyvone/publications_promoottori. pdf> [Accessed 12 April 2005].

Hyvönena, E. Junnila, M., Kettula, S., Mäkelä, E., Saarela, S., Salminen, M., Syreeni, A., Valo, A. and Viljanen, K. (2004b) *Finish museums on the semantic web: the user's perspective on MuseumFinland.* [Online] Available at: http://www.archimuse.com/mw2004/papers/ hyvonen/hyvonen.html [Accessed 12 April 2005].

Ingwersen, P. (1992) *Information retrieval interaction.* London: Taylor Graham

International Standards Organisation (1999) *ISO/IEC 13250 Topic maps.* [Online] Available at: http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0129.pdf

Iter, Inc. (2005) *Iter: gateway to the Middle Ages and Renaissance* [Online] Available at: <http://www.itergateway.org> [Accessed 4 April 2005].

Johnson, E. D. and Harris, M. H. (1976) *History of Libraries in the Western World.* Metuchen, N.J.: Scarecrow Press.

Joint Steering Committee (2004) *Joint Steering Committee for revision of Anglo-American Cataloguing Rules. A brief history of AACR: early English language cataloguing codes.* [Online] Available at: <http://www.collectionscanda.ca/jsc/history.html> [Accessed 13 April 2004].

Katz, W.A. (1997) Reference book. *In:* Feather, J and Sturges, P (eds.) *International encyclopedia of information and library science.* London: Routledge, pp.195-397.

Klein, J. T. (1996) 'Interdisciplinary needs: the current context' *Library Trends* 45(2). 134-54.

Knowledge Management Connection (2004) *Faceted classification of information.* [Online] Available at: < http://www.kmconnection.com/DOC100100.htm> [Accessed 22 March 2005].

Koch, T., Day, M., Brümmer, A., Hiom, D. Peereboom, M. Poulter, A. and Worsfold, E. (1997) *The role of classification schemes in Internet resource description and discovery: Desire project deliverable.* [Online] Available at: < http://www.ukoln.ac.uk /metadata/desire/classification/> [Accessed 18 January 2005].

Koivunen, M. R. and Miller, E. (2001) *W3C Semantic Web activity.* [Online] Available at:< http://www.w3.org/2001/12/semweb-fin/w3csw> [Accessed 30 June, 2006}.

Kop, S. G. and Na, J. C. (2006) 'Semantic relations in information science.' *In:* Cronin, B. (ed.) *Annual Review of Information Science and Technology*, volume 40. Medford, N.J.: Information Today, Inc.

Kuhlthau, C. C. (1993) *Seeking meaning: a process approach to library and information services.* Norwood, NJ: Ablex Publishing Corporation

Kuhlthau, C. C. (1999) 'Accommodating the user's information search process: challenges for information retrieval designers' *Bulletin of the American Society for Information Science* 25(3), 12-16.

Kwasnik, B. H. (1999) 'The role of classification in knowledge representation and discovery.' *Library Trends* 48(1), 22-47.

La Barre, K. (2004) *Adventures in faceted classification: a brave new world or a world of confusion*. Proceedings of the 8th international ISKO conference, 13-16 July 2004. London Conference. [Online] Available at: < http://ella.slis.indiana.edu/ ~klabarre/ isko_04. pdf> [Accessed 18 January 2005].

Langridge, D. W. (1997) 'A new dawn?' *In:* Gilchrist, A. D. (ed.) *From classification to knowledge organization. Dorking Revisited or 'past is prelude'. A collection of reprints to commemorate the 40 year span between the Dorking international conference (first international study conference on classification research, 1957) and the 6th international study conference on classification research (London, UK, 1997).* The Hague: FID, p.x.

Larson, R. R. (1991a) Between Scylla and charybdis: subject searching in the online catalog. Advances in Librarianship 15, pp.175-236.

Larson, R. R. (1991b) The decline of subject searching – long term trends and patterns of index use in an online catalog. *Journal of the American Society for Information Science* 42(3), 197-215.

Lerner, F. (2000) *Libraries through the ages.* New York: Continuum.

Li, J. Z., Gašević, D., Nesbit, J. C. and Richards, G. (2005) *Ontology mappings to enhance interoperability of knowledge domain taxonomies.* [Online] Available at: < http://www.lornet.org/presentation_i2lor_05/papers/i2lor05-07.pdf> [Accessed 30 June 2006].

Library of Congress (2005a) *MARC 21 concise bibliographic: subject access fields (6XX)* [Online] Available at: < http://www.loc.gov/marc/bibliographic/ecbdsubj.html> [Accessed 22 March 2005].

Library of Congress (2005b) *Library of Congress online catalog. Indigenous art: art gallery of Western Austraila. MARC tags.* [Online] Available at: <http://catalog.loc.gov/cgibin/Pwebrecon.cgi?v4=1&ti=1,1&SEQ=20050411065450&Search_Arg=Indigenous+art%3A+art+gallery+of+Western+&Search_Code=TALL&PID=7928&CNT=25&SID=1> [Accessed 22nd March 2005].

Lui, S. (1993) *The automatic decomposition of DDC synthesized numbers.* PhD Thesis, University of California, Los Angeles.

Lui, S. (1996) Decomposing DDC Synthesised numbers. Proceedings of the 62nd IFLA General Conference, 25-31 August 1996, Beijing, China. [Online] Available at: http://www.ifla.org/IV/ifla62/62-sonl.htm [Accessed 22 March 2005]

McArthur, T. (1986) *Worlds of reference: lexicography, learning and language from the clay tablet to the computer.* Cambridge: CUP.

McIlwaine, I. C. (1993) *Guide to the use of UDC: an introductory guide to the use of and application of the Universal Decimal Classification Scheme.* The Hague: International Federation for Information and Documentation.

McIlwaine, I . C. and Williamson, N. (1999) 'International trends in subject analysis research.' Knowledge Organization 26(1), 23-29.

Mäkelä, E., Hyvönen, E. and Sidoroff, T. (2005) View-based user interfaces for information retrieval on the Semantic Web. *Proceedings of the 4th International Semantic Web Conference. Workshop: End User Semantic Web Interaction*, 6-10 November. 2005, Galway, Ireland. [Online] Available at: < http://www.seco.tkk.fi/publications/2005/makela-hyvonen-et-al-view-based-user-2005.pdf> [Accessed 30 June 2006].

Maltby, A. (1975) *Sayers' manual of classification for librarians*, 5th ed. London: André Deutsch.

Marcella, R. and Newton, R. (1994) *A new manual of classification*. Aldershot: Gower.

Marchionini, G. (1995) *Information seeking in electronic environments*. Cambridge: Cambridge University Press.

Markey, K. (1984) *Subject searching in library catalogs. Before and after the introduction of online catalogs*. Dublin, Ohio: OCLC

Markey, K. (1985) Subject searching experiences and needs of online catalog users : implications for library classification. *Library Resources and Technical Services*. 29(1), 34-51.

Markey, K. (1987) Searching and browsing the Dewey Decimal Classification in an online catalog. Cataloguing and Classification Quarterly 7(3), 37-68.

Markey, K. (1989) 'Subject searching strategies for online catalogues through the Dewey Decimal Classification' *In*: Hildreth, C R (ed.) *The online catalogue: developments and directions*. London: Library Association, pp.61-83.

Markey, K. and Deymeyer, A.N. (1984) Dewey Decimal Classification online project: evaluation of a library schedule and index integrated into the subject capabilities of an online catalog. Final report to the Council of Library Resources, report number OCLC/OPR/RR-86/1. Dublin, Ohio: OCLC.

Metcalf, J. (1983) Subject classifying and indexing of libraries and literature. London: Angus and Robertson.

Miksa, F. L. (1994) Classification *In:* Wiegand, W.A. and Davis, D.G. (eds.) *Encyclopedia of library history.* London: Garland, pp.144-153

Miksa, F. L. (1998) *The DDC, the universe of knowledge, and the post-modern library.* Albany, New York: Forest Press.

Miles, A. J. (2005) Announcing SKOS Core 2nd public working drafts. *Faceted Classification Discussion,* 30 November [Online] Available at: < http://article.gmane.org/gmane.comp.infodesign.facetedclassification/556> {Accessed 30 June 2006].

Miles, A., Matthews, B., Beckett, D., Brickley, D. and Wilson, M. (2005) SKOS: a language to describe simple knowledge structures for the web. [Online] Available at: < http://www.idealliance.org/xtech/05/call/xmlpapers/03-04-01.1588/.03-04-01.html#S4.3> [Accessed 30 June 2006].

Mills, J. (1960) *A modern outline of library classification.* London: Chapman & Hall

Mills, J. (1997) Introductory address. *In: Knowledge Organization for information retrieval, proceedings of the sixth international study conference on classification research, University College London, 16-18 June 1997.* The Hague, Netherlands: FID, pp.1-11.

Mitchell, J. S. (1997) Challenges facing classification systems: a Dewey case study. *In: Knowledge organisation for information retrieval, proceedings of the sixth international study conference on classification research, University College London, 16-18 June 1997.* The Hague: FID, pp.85-89.

Mitchell, J. S. (2001) Dewey Decimal Classification : 125 and still growing. *OCLC Newsletter,* November/December.

Netscape (2002) *About the open directory project.* [Online] Available at: http://dmoz.org/about.html [Accessed 22 March 2005].

Netscape (2004) *dmoz open directory project.* [Online] Available at: http://dmoz.org/ (Accessed 22 March 2005)

O'Brien, A. (1994) 'Online catalogs: enhancements and developments' *In:* Williams, M.E. *(ed.) Annual review of information science and technology* 29. Medford, New Jersey: Learned Information, pp.219-242.

OCLC (1998) *Dewey for Windows.* [CD-ROM] New York: OCLC, Forest Press.

OCLC (2004) *Dewey is the world's most widely used classification system.* [Online] Available at: < http://www.oclc.org/dewey/about/translations/default.htm> [Accessed 22 March 2005].

OCLC (2004a) *Using OCLC WebDewey: an OCLC tutorial.* [Online] Available at:< http://www.oclc.org/dewey/resources/tutorial/> [Accessed 1 March 2005].

OCLC (2004b) *Mapping new LCSH with DDC numbers.* [Online] Available at: < http://www.oclc.org/dewey/updates/numbers/> [Accessed 1 March 2005].

Oh, D. G. and Yeo, J. S. (2001) Suggesting an option for DDC class religion (200) for nations in which religious diversity predominates. *Knowledge Organization* 28(2), 75-84.

Olson, H. A. (2001) 'Sameness and difference: a cultural foundation of classification' *Library Resources and Technical Services* 45(3), 115-122.

Passin, T. B. (2005) *Explorer's guide to the Semantic Web.* [Online] Available at: < http://www.manning-source.com/books/passin/passin_chp1.pdf> [Accessed 30 June 2006]

Pepper, S. (2002) *The TAO of topic maps: finding the way in the age of infoglut.* [Online] Available at: http://www.ontopia.net/topicmaps/materials/tao.html [Accessed 22nd March 2005].

Pollitt, A. S. (1997) Interactive information retrieval based on faceted classification using views. *In: Knowledge organisation for information retrieval, proceedings of the sixth international study conference on classification research, University College London, 16-18 June 1997.* The Hague: FID, pp.51-56.

Pollitt, A. S. (1998) 'The application of Dewey Classification in a View-based Searching OPAC' *In:* Mustafa el Hadi et al (eds.) *Structures and relations in knowledge organisation: proceedings of the 5th International ISKO Conference, Lille, France, August 25-29, 1998.* Wurzberg: Ergon Verlag, pp.176-183.

Pollitt, A. S., Braekevelt, P. A. J., Lockwood, H., Tinker, A. J. and Woods, P. (forthcoming). View-based Searching and an Internet database browser. *Communications of the ACM.*

Powell, A. (2005) 'A personal perspective on knowledge management' *In: Knowledge Management Workshop, September, 2005, Nottingham, UK.* [Online] Available at: < http://www.jisc.ac.uk/uploaded_documents/5-%20ukoln-km-perspective-andy-powell.ppt> [Accessed 30 June 2006].

Preece, J., Rogers, Y. and Sharp, H. (2002) Interaction design: beyond human-computer interaction. New York: Wiley & Sons.

PSigate (2004) Psigate cataloguing rules. [Online] Available at: < http://www.psigate.ac.uk/content4.doc> [Accessed 22 March 2005].

Ranganthan, S. R. (1965) *A descriptive account of Colon Classification.* Bangalore: Sarada Ranganathan Endowment for Library Science [reprinted 1990].

RDN (2004) *Resource Discover Network.* [Online] Available at: < http://www.rdn.ac.uk> [Accessed 22 March 2005].

Riesthuis, G. J. A. (1997) Decomposition of complex UDC notations *In: Knowledge organisation for information retrieval, proceedings of the sixth international study conference on classification research, University College London, 16-18 June 1997.* The Hague: FID, 139-143.

Rowley, J. E. and Slack, F. (1998) *Designing public access systems.* Aldershot: Gower.

Satija, M. P. (1997) The revision and future of Colon Classification. *Knowledge Organization* 24(1), 18-23.

School of Library, Archive and Information Studies, UCL (2003) *Facet analytic theory in managing knowledge structure for the humanities.* [Online] Available at: < http://www.ucl.ac.uk/fatks/> [Accessed 4 April, 2005].

Searing, S. E. (1996) 'Meeting the information needs of interdisciplinary scholars: issues for administrators of large university libraries.' *Library Trends* 45(2), pp 315-342.

Semantic Computing Research Group (2004) MuseumFinland: Finish museums on the semantic web. [Online] Available at: < http://museosuomi.cs.helsinki.fi/> [Accessed 12 April 2005].

Sharp, H. A. (1948) *Cataloguing: a textbook for use in libraries.* 4th ed. London: Grafton & Co.

Singh, S. (1999) Potential limitations and the future of the Colon Classification. *Herald of Library Science*, 38(3-4), 186-194.

Smith, I. A. (1997) 'Catalogues' *In:* Feather, J and Sturges, P (eds.) *International encyclopedia of information and library science.* London: Routledge, pp.49-50.

Smith, P.J., Krawczak, D., Shute, S. J. and Chignell, M.H. (1987) 'Bibliographic information retrieval systems: increasing cognitive compatibility.' *Information Services & Use* 7, 95-102.

Soergel, D. (1999) The rise of ontologies or the reinvention of classification. Journal of the American Society for Information Science 50(12), pp. 119-1120 [Online] Available at: < http://www.dsoergel.com/cv/B70.pdf> [Accessed 30 June 2006].

Solomon, P. (2003) 'Discovering information in context.' *In*: Cronin, B (ed.) Annual Review of Information, Science and Technology. Medford, NJ: Information Today.

Suchman, L. A. (1987) *Plans and situated actions: the problems of human machine communication.* Cambridge: Cambridge University Press.

Svenious, E. (1983) 'Use of Classfication in Online Retrieval' *Library Resources and Technical Services* 27(1), 76-80.

Svenious, E. (1992) 'Classification: prospects, problems and possibilities' *In:* N J Williamson & M Hudon (eds.) *Classification research for knowledge representation and organization: Proceedings of the fifth international study conference on classification research.* London: Elsevier Science, pp.5-25

Swanson, D. R. (1964) 'Dialogues with a catalog' *Library Quarterly* 34, 113-125.

SWED (n.d.) *SWED: Semantic Web Environmental Directory*. [Online] Available at: < http://www.swed.org.uk/> [Accessed 30 June 2006].

Sweeny, R. (1983) 'The development of the Dewey Decimal Classification' *Journal of Documentation* 39(3), 192-205.

Sweeny, R. (1990) 'The Atlantic divide: classification outside the United States.' *In:* Bengtson, B G & J S Hill (eds.) *Classification of library materials: current and future potential for providing access*. London: Neal- Schuman, pp. 40-51.

Taylor, R. S. (1968) 'Question-negotiation and information seeking in libraries' *College and Research Libraries* 29, 178-189.

Thompson, J. (1977) *A history of the principles of librarianship*. London: Bingley.

Tinker, A. J., Pollitt, A. S., O'Brien, A., Brakevelt, P. A. (1999) The Dewey Decimal Classification and the transition from physical to electronic knowledge organisation. *Knowledge Organization* 26(2), 80-96.

TopicMaps.org (2001) XML *topic maps (XTM) 1.0: TopicMaps.Org specification.* [Online] Available at: < http://www.topicmaps.org/xtm/1.0/> [Accessed 30 June 2006].

Travis, I. (1982) Faceted classification in an online environment. *Studien zur klassifikation* 11, 269-276.

Treglown, M., Pollitt, A. S., Smith, M. P., Braekevelt, P. A. J. and Finlay, J. E. (1997) *HIBROWSE for bibliographic databases: a study of the application of usability techniques in view-based searching.* British Library and Innovation Centre Report 52. Huddersfield: University of Huddersfield.

Trotter, R. (1993) Dewey and Britain: a continuing partnership. British input to the Dewey Decimal Classification over the past 25 years. *In:*Byford, J, Trickey, K V and

Woodhouse, S (eds.) *AACR2, DDC, MARC and friends: the role of CIG in bibliographic control.* London: LA publishing, pp.61-76.

Trotter, R. (2000) RE: Dewey Table 2 structure [Personal Communication, 5 October 2000].

UC Berkley (n.d.) *The Flamenco search interface project.* [Online] Available at: < http://bailando.sims.berkeley.edu/flamenco.html> [Accessed 21 February 2005].

UDC Consortium (2002) About the Universal Decimal Classification and the UDC Consortium. [Online] Available at: < http://www.udcc.org/about.htm> [Accessed 22 March, 2005].

Vdovjak, R., Barna, P. Houben, G. J. and Frasincar, F. (2003) *Bringing the Semantic Web closer to the user.* [Online] Available at: http://www2003.org/cdrom/papers /poster/p093/p93-vdovjak.htm [Accessed 30 June 2006].

Vickery, B. C. (1960) *Faceted classification: a guide to the construction and use of special schemes.* London: Aslib.

View-Based Systems Ltd. (2004) *View-based systems: improving access to information and knowledge.* [Online] Available at: < http://www.view-based-systems.com/>

Vizine-Goetz, D. (1997) 'OCLC investigates using classification tools to organize Internet data' OCLC Newsletter 226, March/April [Online] Available at: < http://www.oclc.org/oclc/new/n226/research.htm> [Accessed 25 June 1998].

Vizine-Goetz, D. (1998) Dewey as an Internet subject guide. *In:* Mustafa el Hadi et al (eds.) *Structures and relations in knowledge organisation: proceedings of the 5th International ISKO Conference, Lille, France, August 25-29, 1998.* Wurzberg: Ergon Verlag, pp.191-197.

Vizine-Goetz, D. (2002) Classification schemes for Internet resources revisited. *Journal of Internet Cataloguing* 5(4) [Online] Available at: < http://staff.oclc.org/~vizine/ JIC/v5n42002/ClassificationSchemesRevisited.doc> [Accessed 22 March 2005].

Wajenberg, A. S. (1983) MARC coding of DDC for subject retrieval. *Information Technology and Libraries*, September, 246-251.

Walker, S. (1991) 'Subject access in online catalogues.' *OPACs and the user: Proceedings of the third Anglo-Nordic Seminar 8-11 April, 1990. British Library R & D Report 6040.* [London]: The British Library Board, pp.23-32.

Wellisch, H. H. and Smiralia, R. P. (1993) Classification. *In*: Wedgeworth. R. (ed.) *World encyclopedia of library and information services*, 3rd ed. Chicago: American Library Association, pp.207-212.

Williamson, N. (1998) 'An interdisciplinary world and discipline based classification'. In: W M el Hadi, J Maniez and A S Pollitt (eds.) *Structures and relations in knowledge organisation: proceedings of the fifth international ISKO conference 25-29 August, Lille, France.* Wurzburg: Ergon Verlag, pp.116-124.

Wilson, R. (2004) *The role of ontologies in teaching and learning.* [Online] Available at: < http://www.jisc.ac.uk/uploaded_documents/ACF11AC.pdf>. [Accessed 30 June 2006]

Wilson, T. D. (1999) 'Models in information behaviour research' *Journal of Documentation* 55(3), 249-270.

World Wide Web Consortium (2004a) *RDF primer.* [Online] Available at: < http://www.w3.org/TR/rdf-primer/> [Accessed 30 June 2006].

World Wide Web Consortium (2004b) *RDF Vocabulary Description Language 1.0: RDF Schema.* [Online] Available at: < http://www.w3.org/TR/rdf-schema/> [Accessed 30 June 2006].

World Wide Web Consortium (2004c) *SKOS Mapping vocabulary specification.* [Online] Available at:< http://www.w3.org/2004/02/skos/mapping/spec/> [Accessed 30 June 2006]

World Wide Web Consortium (2005) *SKOS Core guide.* [Online] Available at: < http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20050510/> [Accessed 30 June 2006].

World Wide Web Consortium (2006a) *WC3: World Wide Web Consortium: leading the Web to its full potential.* [Online] Available at: < http://www.w3.org/>. [Accessed 30 June 2006].

World Wide Web Consortium (2006b) *Extensible Markup Language (XML).* [Online] Available at: < http://www.w3.org/XML/> [Accessed 30 June 2006].

World Wide Web Consortium (2006c) *A survey of RDF/Topic Maps interoperability proposals.* [Online] Available at: < http://www.w3.org/TR/rdftm-survey/> [Accessed 30 June 2006].

Xie, H. (2000) 'Shifts of interactive intentions and information-seeking strategies in interactive information retrieval' *Journal of the American Society for Information Science* 51(9), 841-857.

Yee, M. M. and Layne, S. S. (1998) *Improving online public access catalogs.* Chicago, London: ALA.

Yee, M. M. (2003) *Guidelines for OPAC displays, prepared for the IFLA Task Force on Guidelines for OPAC Displays, September 30, 2003, draft for worldwide review* [Online] Available at: <http://www.ifla.org/VII/s13/guide/opacguide03.pdf> [Accessed 22 March 2005].

Yee, K., Swearingen, K., Li, K. and Hearst, M (2003) 'Faceted metadata for image search and browsing.' *Proceedings of the ACM CHI Conference, 5-10 April, 2003, Ft Lauderdale, Florida, USA.* [Online]< http://bailando.sims.berkeley.edu/papers/flamenco-chi03.pdf> [Accessed 11 January 2005].

Appendix 1

A Report of an Exploratory OPAC-User Study at the University of Huddersfield Library

This pilot study focuses on OPAC use in the real world setting of the Central Library at the University of Huddersfield. It was intended to provide an insight into the information needs of, types of search performed and the nature of any problems experienced by students of this particular university, and it will form a basis for more formal research.

1. Study domain

The University of Huddersfield Library is situated in three locations: the Central Library (serving the Schools of Applied Sciences, Computing and Mathematics, Design Technology, Engineering, Business, Human and Health Sciences, Humanities), the Music Library, and the Holly Bank Library (serving the School of Educational and Professional Development). The university has approximately 17,000 full-time, part-time and sandwich course students.

The library OPAC is Ameritech's Horizon system <http://www.als.ameritech.com/product/horizon.htm> which comprises a windows based menu-driven interface with phrase and keyword searching using the following user-specified indexes for which the user formulates a query:

- Title
- Title Keyword
- Author
- Subject Keyword
- Author/Title Keyword
- Journal Titles
- Shortloan items, arranged by lecturer
- Audio-Visual titles
- Subject
- Series
- Author Keyword
- ISBN
- Dewey Classification numbers
- Music Classification numbers
- E-journals and e-texts
- Theses/dissertations classified search

Users have the option to perform multi-index searches within the four keyword fields, and restrict a search by type of material and/or library location.

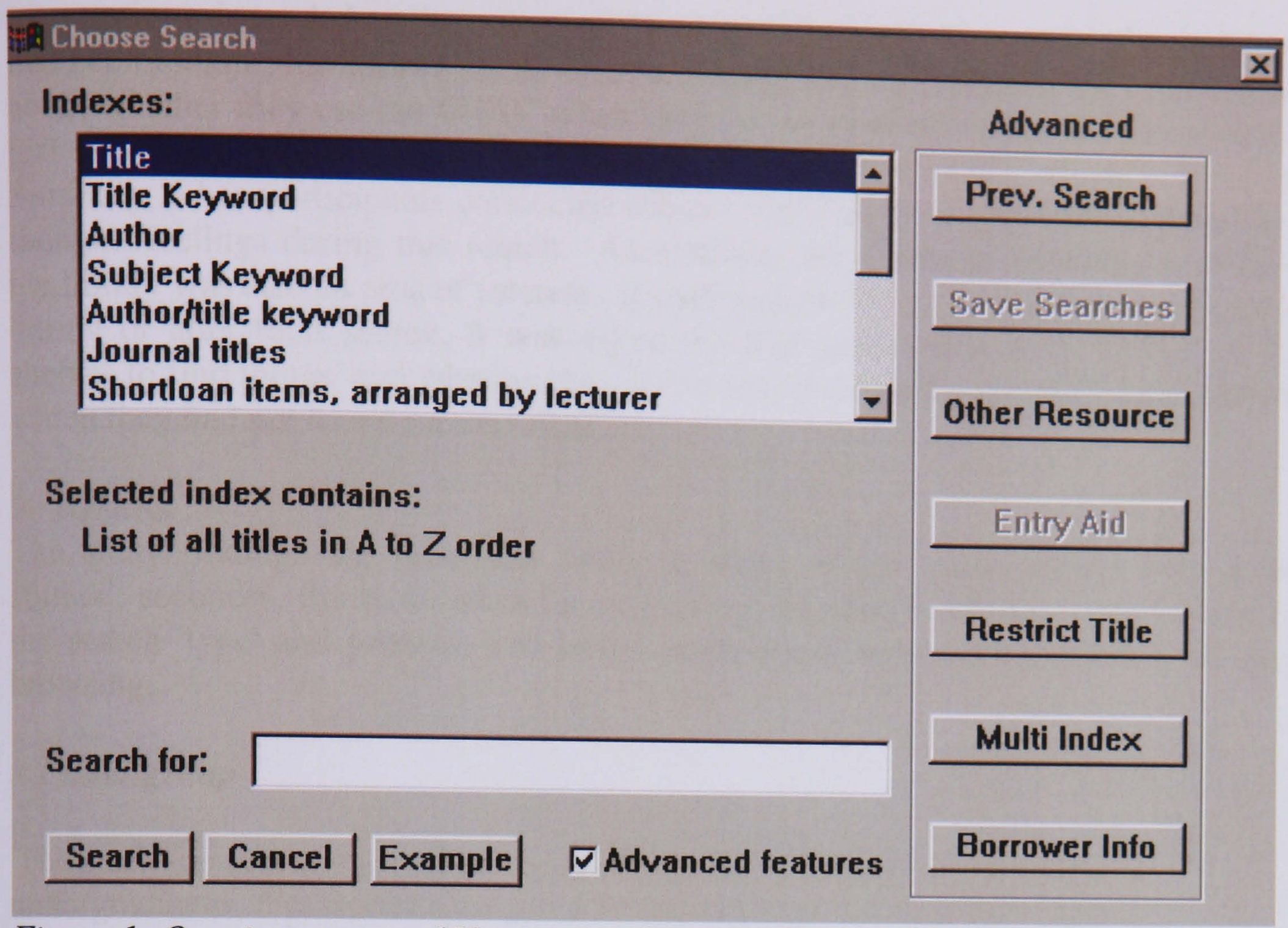


Figure 1: Opening screen of University of Huddersfield OPAC

A Web-based interface is also available for remote access from users' own desktops.

3. Method

Data was collected during 9/10 December 1999 and week beginning 13 December 1999, the busy end of term period. The sampling strategy was non-probabilistic with the researcher approaching people using the OPACs in the library entrance hall during these particular dates. Sixty-nine people were approached and only one refused to participate, resulting in a study sample of 68 participants. Since this study was preliminary, aimed at getting a feeling of the issues involved to guide future research with this user population, generalisability was not a prime concern. By sampling in this way, we can observe OPAC-user interactions that are motivated by real information needs and problems.

The data collection technique combined semi-structured interview with observation of the OPAC search. Data was recorded on observation sheets for each participant. Participants were asked about:

- The information they were looking for
- Their reason for doing the search

Brief demographic information (course/pathway, year of study, sex) was recorded. The search was then observed; participants were reassured that this was not to judge how well they performed the search but simply to see how they looked for information. The search process was recorded on the observation sheets, noting the indexes used (e.g. author, title, title keyword etc.), the participant's search statements and any artefacts used (e.g. recommended reading list, hand written notes). Follow-

up questions depended on the nature of the observed search. Generally, if participants had been looking for 'known-items' from a reading list, for example, they were asked about whether they use the OPAC when they do not have title or author information, just a general subject area, and to describe their experiences of such searches. Similarly, when participants conducted subject searches, they were asked about their thoughts/feelings during this search. Ascertaining the extent of browsing activity in the library was also an area of interest. If participants did not mention this unsolicited during or after their search, it was asked whether participants ever 'looked on the shelves to find things' and whether they found useful things by looking on the shelves which they had not found on the OPAC.

3. Results

The study findings are presented firstly in terms of the nature of the user group studied; secondly, the motivation for instigating the search; thirdly, observations on the search 'type' and process; and lastly, participants' reported experiences of shelf browsing.

3.1 User group

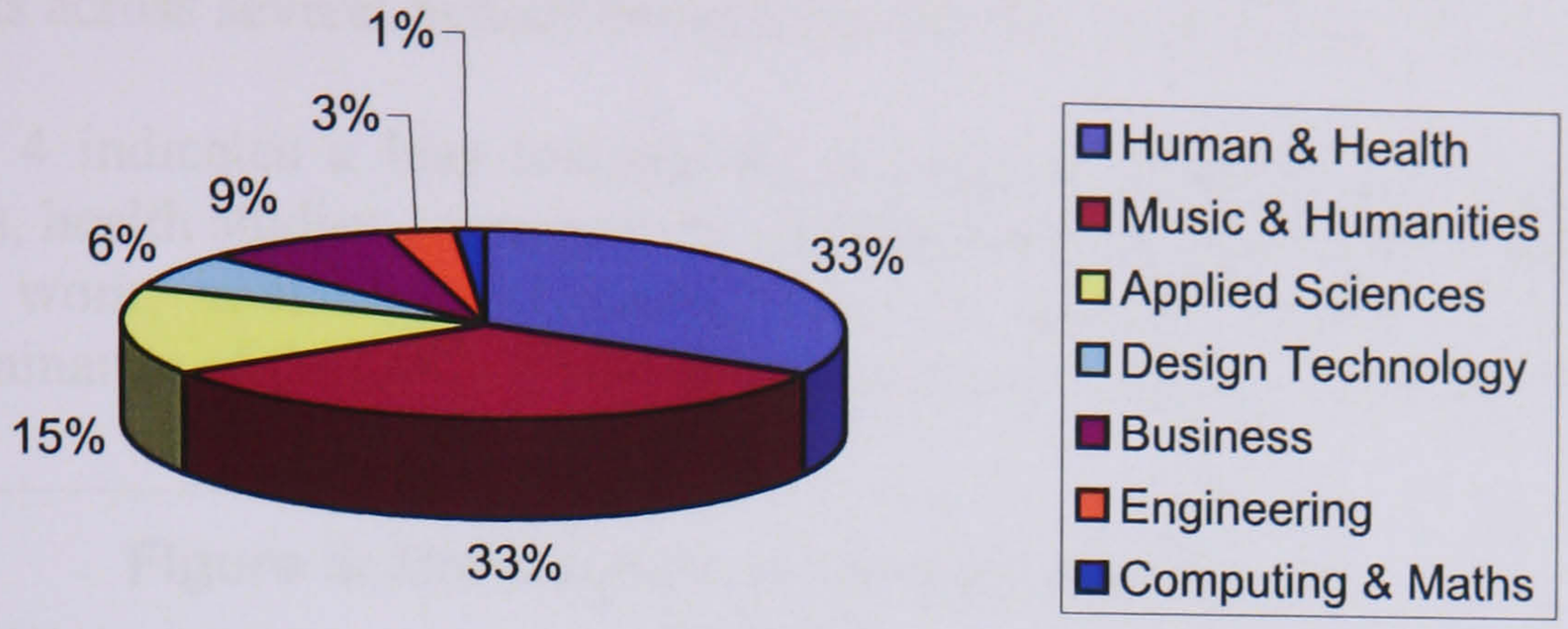
The study sample comprised 68 users (45 female, 23 male), and included 60 undergraduates, 6 postgraduates and 1 lecturer (1 external student's status is unknown) with a roughly equal split between arts and science awards:

Undergraduate		**Postgraduate**		**Other**	
Award	No. of participants	Award	No. of participants	Status	No. of participants
BA	25	MA	1	Lecturer	1
BSc	26	MSc	2	Unknown	1
BEng	2	PgDip	2		
LLB	2	PgCert	1		
DipHe	5				

Figure 2: Academic status of participants

The participants were from all of the seven Schools whose information resources are held at the Central Library. There were no participants from the one remaining School, Education, since their information resources are held in a different library on a separate site:

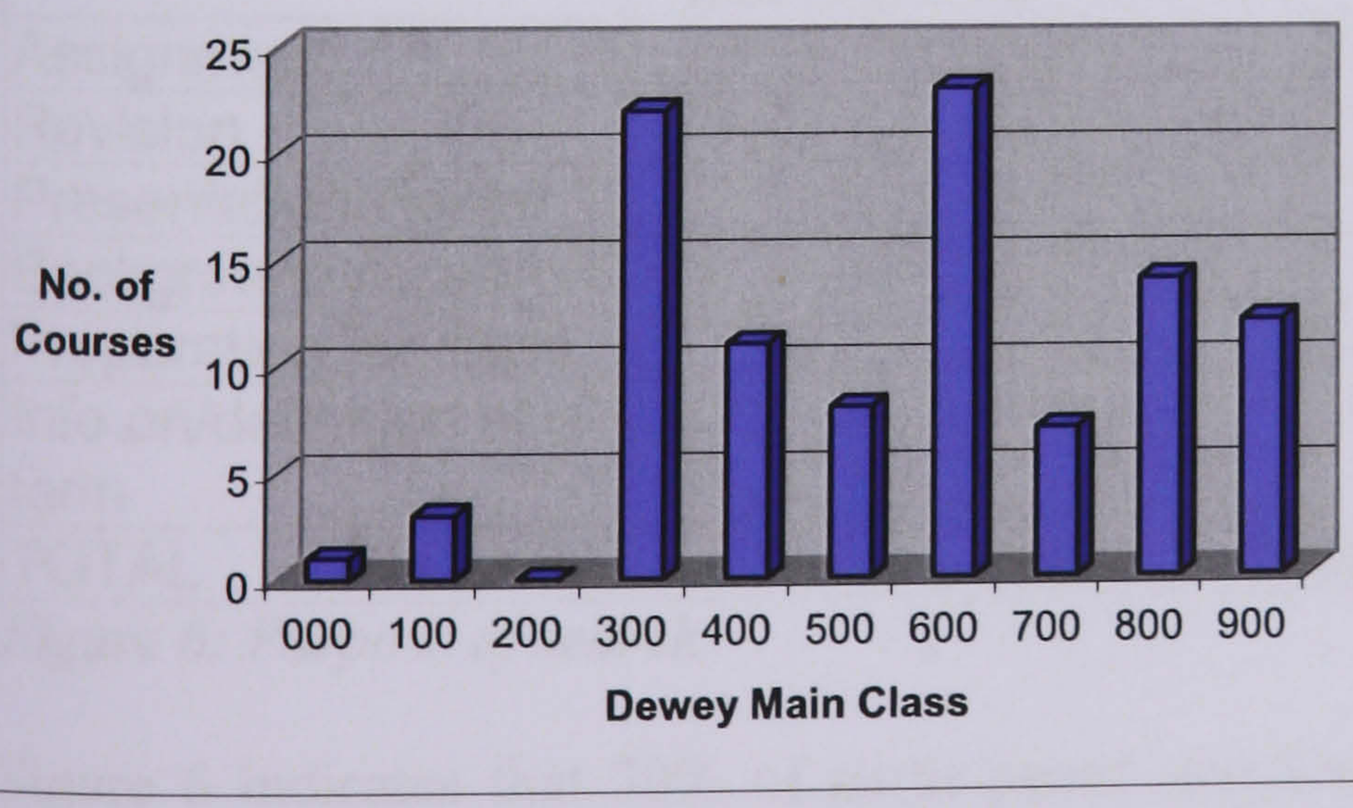
Figure 3: OPAC use by School



As can been seen from figure 3, the Schools of 'Human and Health Sciences' and 'Music and Humanities' were highly represented in the study sample. Participants from the School of 'Music and Humanities', however, were all from humanities disciplines as music students have their own separate library within the music department building. Users from the School of Computing and Maths were notably underrepresented. Only one student from this School was interviewed, who, on this his first time using the OPAC, was then observed searching for books from the subject discipline of education on behalf of another student.

Another way of viewing this subject of study data is, for example, to code it by the main classes of a classification scheme such as Dewey:

Figure 4: Participant Courses by Dewey Class

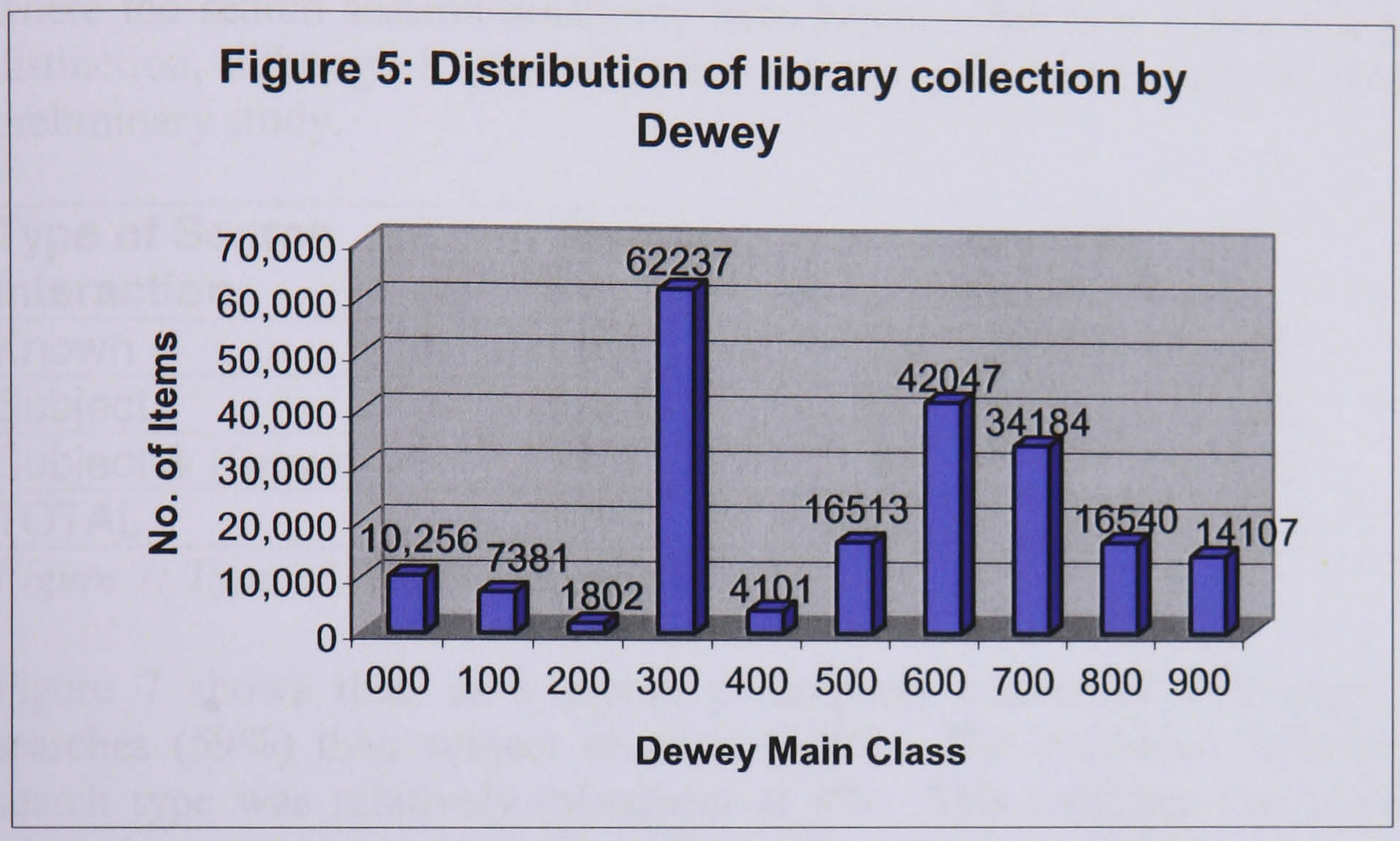


Key
000 Generalities
100 Philosophy & Psychology
200 Religion
300 Social Sciences
400 Language
500 Natural Sciences & Maths
600 Technology (Applied Sciences)
700 The arts fine and decorative arts
800 Literature and rhetoric
900 Geography & History

Figure 4, shows the distribution of participants' courses according to the Dewey Decimal Classification Scheme. The majority of courses were classified to one main class. However some, such as 'English Studies' (language 400 and literature 800) and 'Health and Sports Studies' (600 and 700), were given two or more codes. Figure 4 provides a more finely grained and slightly different characterisation than viewing the

course data by school (figure 3). For example, 'Applied Sciences' in figure 3 includes subjects such as biology, chemistry, whereas in Dewey these would be classified as '500 Natural Sciences…', and '600 Technology (Applied Sciences)' would include subjects across several School boundaries (Health, Engineering, Design Technology).

Figure 4 indicates a bias towards the disciplines of 600 technology (e.g. nursing, textiles, health studies, management, engineering) and 300 social sciences (e.g. media, social work, sociology, economics) in the sample studied. Interestingly, the predominance of these two fields is reflected in the library collection itself:



3.2 Purpose of OPAC search

Participants' reasons for searching were all, to varying degrees, task-driven.

Purpose	No. of participants	%
Assignment	47	70
Revision	9	13
Presentation/Poster	5	7
Background research	3	4
Preparation for class	2	3
Info on/definition of a term	2	3
TOTAL	68	100

Figure 6: Purpose of search

Figure 6 indicates that 70% of participants' searching was externally motivated to fulfil the requirements of an assignment. This is unsurprising, given the time of year of the investigation. Similarly, assessed presentations and exam revision also provided impetus for searching. A slightly less task-driven reason for searching was background reading, namely for a dissertation, work placement and about an author of interest. Two participants were preparing for class - to give a lecture and attend a tutorial. Lastly, two participants were seeking information on/definition of a specific term. One student was searching for an encyclopaedia of textiles to find information

about 'lace' (presumably for an assignment) whilst another (nursing) student sought a definition and prescribed treatment for a particular medical condition with which a family member had recently been diagnosed.

3.3 The search interaction

The degree of information known to the participant about the search item was used during data analysis to classify the observed searches as 'Known', 'Subject' or 'Known and Subject'. 'Known item' searches are those where the participant had information for retrieving a specific author or title, 'Subject' searches are those where the participant had no author or title information, and 'Known and Subject' are those where the search session combined both types. This is a classic but perhaps crude distinction, although it proved useful for the general observation purposes of this preliminary study.

Type of Search Interaction	**Number**	**%**
Known	40	59
Subject	22	32
Subject & Known	6	9
TOTAL	68	100

Figure 7: Type of search interaction

Figure 7 shows that, as a group, participants conducted 27% more known item searches (59%) than subject searches (32%). The combined 'Subject and Known' search type was relatively infrequent at 9%. This indicates that participants more often than not had a clear idea about what they were looking for. The observed interactions are now discussed in the framework of these search types.

3.3.1 Known-item searching

The predominance of known-item searching observed in the investigation was expected owing to the fact that the student participants were all taking taught courses. Twenty-nine out of the 40 known-item searches were aimed at retrieving items recommended by a tutor. Participants were armed with reading lists in the form of course booklets, assignment instructions and hand written notes to aid their search. Other sources for these 'known items' were verbal recommendation by tutor, recommendation by a friend, results of CD-ROM search, previous OPAC searches, journal citations and book bibliographies.

Index used for known items	No. of participants
Author	33
Title	25
Title Keyword	8
Journal Title	6
Author/Title Keyword	2
Dewey Code	1
ISBN	1

Figure 8: Indexes used by participants for Known-item searches

Figure 8 shows the relative take-up of search options for known items according to the number of participants (not individual searches) selecting these. It indicates that 'Author' searches, closely followed by 'Title', were the most popular choice by participants for finding known items. In these types of search, participants were generally successful in finding the sought-after item; however, even with accurately written references problems still arose.

Index	Successful	Search Failure	Collection Failure	Number
Author	18	33	4	55
Title	20	12	16	48
Journal Title	6	5	4	15
Title Keyword	2	4		6
Author/Title Keyword		2	1	2
ISBN	1		1	2
Dewey Code	1			1
TOTAL	48	56	26	129
%	37	43	20	100

Figure 9: Matching rate of known item searches

Figure 9 shows instances of 'successful' and 'unsuccessful' searches. A search was coded 'successful' when the user retrieved the sought after item directly after inputting the query, i.e. entailing no scrolling to find the item and/or iteration. A search failure occurred when the user query failed to match the system indexing language; whilst a collection failure resulted in no hits when the item was not held by the library. As figure 9 indicates, only 37% of known item searches were initially successful, leaving a 63% failure rate (43% Search Failure, 20% Collection Failure). It can also be seen that searches for known authors were most frequent; however, less successful at 18 (33% of author searches) than known item title searches at 20 successes (42% of title searches). Notably, all the predominantly used index types (exceptions being ISBN and Dewey Code) resulted in more failures (combined search and collection failure rates) than successes. This is quite a surprising finding, considering that participants typically held reasonably comprehensive and reliable information (taken from assignment instructions and module reading lists) about the sought-after items. Participants were observed experiencing problems in transforming this information into the appropriate system language during the OPAC interaction.

Failed searches occurred when the search statement was entered in the wrong format. For example in the case of 'author', one participant entered two author names ('Buchanan and Luczynki') in a single author field, another entered the full author name without inversion ('Anna Chesner', and another inverted the author name but did not separate the first and last names with a comma ('Hill Christopher'). Author names such as 'De Thornyay' and 'De Shazer' also resulted in initial search failure when two participants entered these as a single word; both were correctly entered on a second attempt. Failed searches also occurred when hand-written author references were misspelled. For example, one participant searched for 'Levis' instead of 'Leavis', another 'Morley, David' then successfully 'Morely, David'. One participant was unsure of the order of the author's names "Neal Zora Hurston or Hurston Zora Neale?", by trial and error the latter was discovered to be correct.

Title searches were less problematic than those for authors, but the format caused problems for one participant who prefaced it with an edition statement ('Fourth edition sociology themes and perspectives'). Fortunately, though, she then searched for the same item by author ('Haralambos') with successful results. Such persistence was often observed if an initial search failed and the alternative author or title information was also available for a second attempt. Another known title failure search was observed when two students, working co-operatively, searched for journal *article* titles in the *journal title* field. They also experienced problems when searching for journal acronyms rather than full titles.

Twenty-nine (53%) out of 55 known author search statements were on the author's surname alone. This caused little problem when the name was unusual but common names often involved the user in time-consuming list scrolling to find the required author. This was compounded when only the author surname or surname and initial was known to the user, resulting in individual inspection of each author entry until the desired title was found, or abandonment and instigation of an alternative title search strategy.

3.3.2 Subject Searching

Figure 7 showed that 23 (33.8%) participants engaged in subject searching. These students often had no recommended titles to fulfil their specific information requirement. This was often because the assignment, particularly in the case of oral presentations, required students to choose their own topic within a subject area. Three students subject searched by default as they had forgotten to bring or lost their reading lists.

Index used for subject search	No. of participants
Subject Keyword	20
Title Keyword	12
Title	6
Author/Title Keyword	4
Subject	2
Author	1
Short loan (by lecturer)	1

Figure 10: Indexes used by participants for subject searches

It can be seen from figure 8 that 'Subject Keyword' was the most popular choice of index for subject searches. Interestingly, 6 participants searched for a subject using the title field, typically used for known items; the 1 participant searching by 'Author' was actually seeking critiques of Arthur Morrison, but employed an author search strategy to obtain a Dewey code after a failed 'Subject Keyword' search.

In general, subject searching proved more problematic than known item searching. Participants tended to enter a subject term which was either too broad, leading to an unmanageable number of hits, or too narrow, resulting in zero hits. For example, one student was searching for information about 'linguistics' for an assignment on 'regional variation and pronunciation'. He entered the term 'linguistics' in the subject keyword field, which resulted in 174 subject heading hits. He scrolled down the first few headings and noted down Dewey class codes, describing the OPAC as 'a vague pointer...if you know what book you're looking for it's good'. He then explained that he would 'go look at these class numbers and around the general area'. Similarly, another student seeking items on 'manufacturing and enterprise' did a Subject Keyword search on 'Manufacturing', which resulted in 573 hits. He chose one title and noted the class number 658.562, saying "I've got a number now so I'll go look on the shelves". This searching to find a class number for shelf browsing was a common strategy if an OPAC subject search failed to retrieve the desired items.

One participant, searching for items on the topic of climate, searched specifically to obtain relevant class numbers. She carried out five title key word searches on 'Climate', 'British Climate', 'British Weather', 'Hydrology in Britain' and 'Hydrology', examined a few titles and noted the class numbers alone each time. She showed me her results: 'I'm noting these down so I'll go look on the shelves. See you're getting a pattern (of similar numbers) emerging here'.

The actual terms used in subject queries are interesting:

Figure 11: Information sought and subject terms used in queries

Information sought	**Subject terms (as used)**
1. Criticism on Arthur Morrison & Charles Darwin	Criticism Morrison, Darwinism, Darwinian theory
2. Linguistics (Essay on regional variation, pronunciation)	Linguistics, Pronunciation
3. Encyclopaedia of textiles	Encyclopedia
4. Rugby	Americanisation, rugby, marketing
5. Drama therapy	Drama therapy
6. Books on prostitution and syphilis	Syphillis, Prostitute
7. Corporate books	Corporate reporting, corporate, corparate, reporting
8. Lace	Lace, Raschel, warp knitting, knitting
9. Dyeing machines	Dyeing machines
10. Deviancy and media effects	media effects, the effects on media, social deviancy, deviancy, media affects on society, deviance in society, sociology
11. Basic chemistry books	Introduction to spectroscopy
12. Criticism on Elizabeth Gaskell and Charles Dickins	Gaskell, Dickins
13. Air pollution, urban geography, social exclusion (for poster presentation)	Presentations, Poster presentations
14. Family care at work	Family care at work, family care
15. Slavery	Slavery
16. Visual culture and technology	Scientific imaging, medical imaging, virtual reality, digital photography (terms supplied by tutor)
17. Investment analysis	Investment
18. Manufacturing and enterprise	Manufacturing
19. Books on pesticides	Pesticides
20. Digital TV, implications of	Digital TV
21. Psychologists, self esteem, assertiveness	Carl Rodgers, reflections in nursing, reflection in nursing, IT and nursing, Computers and nursing, Psychology and self esteem, self concept
22. Climate	Climate, British weather, hydrology in Britain, hydrology
23. Content analysis	Content analysis
24. Haemochromotosis	Haemochromotosis, Hemochromotosis
25. Women in fiction, Radcliffe Hall, Black African stories	Lesbianism and fiction, lesbianism, American African Writing, Black writers

It can be seen from figure 11 that participants' query specification typically contained the terms they had used to describe their information need (information sought), i.e. there was little translation or expansion of the information need as necessary into the 'language' (item representation) of the system. As a result, subject searching was a rather hit and miss affair with users sometimes finding a matching keyword (e.g. 'slavery', 'content analysis', 'encyclopedia' and unambiguous proper nouns such as 'Gaskell' and 'Dickens') and sometimes not. However even when successfully matching a keyword, searches frequently produced a large, often unclassified, result

set through which the user had to scroll, for example, 'sociology' (3663 hits), 'linguistics' (569), 'manufacturing' (573), 'corporate' (599), 'investment' (166), 'climate' (180), 'pesticides' (176).

Users experienced failed searches when the terms were simply not indexed for a particular index field (for example, in the case of 'Darwinism', 'prostitute', 'scientific imaging', 'medical imaging', 'American African Writing' and 'deviancy'). Postmodified noun phrase formulations (e.g. 'the effects on the media', 'media effects on society', deviance in society', 'reflections in nursing', 'family care at work', 'hydrology in Britain') and Boolean AND statements (e.g. 'lesbianism and fiction', 'IT and nursing' 'psychology and self-esteem') also resulted in no hits. In addition, misspelled words ('syphillis', 'corparate', 'Haemochromotosis') or spelling variants (americani*s*ation/ americani*z*ation, encylop*e*dia/encylop***ae***dia) also affected recall or produced no hits.

Despite the limited number of terms provided by users, there appears to be an awareness of hierarchy and the need to broaden and narrow searches as required. For example:

Warp Knitting	<	Knitting
Presentations	>	Poster Presentations
Social deviancy	<	Deviancy
Corporate reporting	<	Reporting

Other participants converted both postmodified noun phrases (by prepositional phrases) and coordinated noun phrases into a noun phrase, which broadened the search by reducing the number of concepts:

Family care at work	<	Family care
Hydrology in Britain	<	Hydrology
Lesbianism and fiction	<	Lesbianism
Deviance in society	<	Sociology

There were examples of synonym/quasi-synonym usage:

Climate	→	British *Weather*
IT and nursing	→	*Computers* and nursing
Self-esteem	→	Self concept
American African writers	→	Black writers
Darwinism	→	Darwinian theory

And transition from pre to post noun phrase modification, i.e. syntactic variation:

Media effects	→	the effects of the media
Social deviancy	→	Deviance in society

When asked about their thoughts on subject searching (i.e. searching without an author or title), comments ranged from 'okay' to 'time consuming', 'awful'. In general, participants felt that searching without an author or title was not specific enough or similarly too general/broad giving you too much, described by one user as 'difficult to

pinpoint. You search on sport and you get tennis, football, everything'. One participant said it was difficult 'when you don't get anything and you have to think of another word' and another 'it's difficult when you don't really know what you're looking for'. Nine participants said that they very rarely or had never used the OPAC for subject searching, they were always looking for something specific (a known-item).

3.4 Browsing

The majority of participants said that they did browse the library shelves and found this useful. As one mature student said: "It's a bit like Tescos. You always end up with something you don't want …but that's not a bad thing". For some this browsing was due to a failed search: "You can't specify, you have to go to the general area and look through a list. Nine times out of ten I'll go and have a look on the shelves". Others went to the shelves with a Dewey code for a single title to browse for other relevant items in that particular area. Looking at the actual item was how many participants made their relevancy judgement; subject headings in OPAC records were largely ignored.

Some participants would browse the shelves from the outset, before even consulting the OPAC: "I sometimes think 'where do I start' and often go and look on the shelves'". A first-year Law student had never used the OPAC for subject searching, explaining that she usually just looked on the shelves without having a Dewey code "because they are all in sections." One first year student who was using the OPAC for the first time said that browsing was "good for when you don't really know what you want. If you know what you want then you can look on the PAC". Another first-time user, a first year Health and Sports Studies student, was asked how he had found books in the past; he replied that he had "gone to the section on sports and nursing and found it there".

4. Discussion

This study has examined OPAC use at the University of Huddersfield Library in terms of the nature of the user group studied (academic status and type of award, School affiliation, subject discipline of course); participants' reasons for searching; the nature of the observed search interaction, following the traditional distinction made between known-item and subject searches; and the reported browsing strategies of participants. This discussion section relates the study findings to a selection from the literature, and considers how a view-based searching interface could contribute in addressing some of the problems experienced by users.

The *user group* studied consisted predominantly of undergraduate students from the disciplines of technology (e.g. nursing, textiles, health studies, management, engineering) and social sciences (e.g. media, social work, sociology, economics) which was found to be a pattern reflected in the nature of the library collection itself. Further study, however, would be required to determine whether this is an accurate reflection of the pattern of OPAC use. The current investigation was carried out in the Entrance Hall of the library, where no subject discipline predominates. A more representative sample might be achieved by combining this with data collection at each of the subject floors. However, this is somewhat problematic due to silence restrictions in these parts of the library.

The users in this study were extremely directed in their *search goals/purpose* and did not seem to invite changes or even development. Both for known-item and subject searching, the focus was on an impending task, usually an assignment. Known-item searches predominated over subject searches. Particularly for known item searches, relevancy seemed to be decided *a priori*, usually by a third party: a lecturer had recommended the item therefore it was highly relevant. In these known-item searches, participants generally seemed reluctant to stray from the recommended readings, even when these items were on loan.

This relates to Hert's (1995) findings from a naturalistic study of OPAC-user interactions in a university library. She discovered that although the search interaction was situated, in contrast to the findings of several researchers whom she cites (viz. Bates, Harter), users' goals remained relatively static. She states that the traditional match paradigm in information retrieval assumes that user goals are static and led to efforts being concentrated on algorithms for improving the match between queries and document representations. The emergence of user-based approaches led to the idea that goals change during the interaction; however, Hert's own findings disagree with this. She claims that this may be due to a lack of distinction in the literature between the terms 'goal' and 'information need', and suggests that the goal of a specific OPAC search may be part of a larger information seeking process motivated by an information need which may change over time (cf. Kulthau,1997). She also suggests that her findings of static goals may be explained by Harter's concept of 'weak relevance', whereby users are unable to determine fully the relevance of an item until they have actually seen it, and thus are reluctant to change their interaction goal on the basis of a mere citation. Several participants in the current study emphasised their need to view items; many were using the OPAC simply to locate a Dewey class code (without noting the details of specific items) in order to pursue their search at the shelves.

In a View-based OPAC the user is presented with the entire library collection, categorised by subject discipline using Dewey captions that can be directly manipulated to broaden or narrow a search. Viewing and searching with these classificatory structures may potentially inform the search process and the user's knowledge state (Mills, 1997). It may be interesting to explore whether or not users would develop their search goals more with a view-based than with a traditional menu-based/query driven interface, and whether this is to desirable ends. Or perhaps only a full-text system would provide sufficient information to warrant any change in goal at the system interface.

The *search interaction* was broadly categorised according to the level of information the user held about the desired item: a known-item (details of author or title known), subject search (no clear item details but topic information), or a combination of the two within a single search session: known-item/subject. These search types are common in the user study literature. Baker and Lancaster (1991) warn that many apparently 'known item' author or title searches may in fact be 'disguised subject searches'. This was circumvented in the present study by taking account of the entire search interaction, questioning the user about the purpose of their search and noting evidence such as reading lists. The type of index (access point) used was also not directly equated with the type of search interaction, for example six participants searched for a subject using the title index.

Known-item searches were surprisingly problematic, resulting in more initial failures than successes. Although it depends on how 'success' is defined, the success rate at 37% is low compared with previous studies, which report success rates between 70-80% (Baker and Lancaster, 1991 p.203). Thus, despite often having both an author and title, participants encountered problems in formulating and matching their query, typically due to misspellings, mistyping and inappropriate input format (search failure); or less frequently, when the item was not held by the library (collection failure). Previous studies also indicate that 10-45% of users bring incorrect or incomplete citations for searching; the success rate is generally lower for memorised information than written citations (ibid.). The current study concurs with the findings of previous studies which show title searches for known items to be more successful than author searches; however, unlike these studies, author searches were found to be performed more frequently than title searches.

The incorporation of spell checkers and stemming, an early example being the prototype OKAPI catalogue in 1984 (Walker, 1989), can help to help minimise search failures in the above circumstances. Another possible solution might be to encourage lecturers to incorporate ISBN numbers into their reading lists, coupled with promoting a greater awareness amongst users about utilising this unique number for searching. Only one participant in the current study employed this very efficient strategy for the retrieval of known items.

In this study, the success of known-item searching was judged by examining the first user input and accompanying system feedback. 'Success' was presented as an either/or situation for these initial stages of the interaction, following the 'one shot' approach of traditional recall/precision metrics. The reality, however, is a continuum of success (or failure) rather than a dichotomy. For example, participants may fail to retrieve an item initially but then succeed after scrolling to the appropriate point in the result set, or after a first re-formulation, second reformulation and so on. Or they may abandon the search completely without having retrieved the desired item in an apparently failed search. However, failure to retrieve the desired item might with hindsight even be perceived as search success when, for example, the search result points to a more useful item than the original one in mind. Similarly, as Large and Beheshti (1997,123) note "A research student seeking a fresh topic to work on may be delighted to discover a virgin field." Success or failure determined technically by the degree of match between user and system language, although important, is not entirely adequate. Future study should incorporate user perceptions of success.

Subject searching was found to be problematic for many participants in the study, a finding well documented in the literature (O'Brien, 1994; Walker, 1991; Markey, 1989). Markey (1989) found that "between one-third and one-half of the subject terms that users enter fail to retrieve bibliographic records in the online catalogue" (p.82), with the converse problem in 30-40% of cases of too many hits (Walker, 1991). Common problems cited in the literature, all of which were observed in the present study to varying degrees, include:

- Formulating a query containing the 'right' (matching) subject heading
- Entering search terms which are too specific or too general
- Broadening the search if too little is retrieved

- Typographical errors and misspellings

It was found that participants generally formulated their initial query containing *search terms* from their original verbal expression of the information they sought, and if this failed their alternatives were limited in number and semantic scope (e.g. typically a morphological or syntactic variation of the initial formulation). This could perhaps be interpreted as a lack of awareness of controlled vocabulary amongst these participants.

It relates to Taylor's (1968) classic and much cited paper based on reference desk consultations which recognised four levels of information need, forming part of a question-negotiation process continuum:

> Q1 - the actual, but unexpressed need for information (the visceral need)
> Q2 - the conscious, within brain description of the need (the conscious need)
> Q3 - the formal statement of the need (the formalized need)
> Q4 - the question as presented to the information system (the compromised need).
>
> Taylor (1968, 182)

Taylor states that the question expressing the 'formalized need' should be "recast in anticipation of what the files can deliver" thus forming 'the compromised need(s)'..."so that the total system can be searched efficiently". The participants in the current study seemed to remain essentially at the 'formalised need' phase by expressing their query using this same terminology and little beyond this. Variation tended to be morphological or syntactic rather than semantic. Similarly, Tagliacozzo (cited in Baker and Lancaster, 1991) found that 80% of search terms were included in users' original expression of their information need. This lack of significant query transformations perhaps suggests that users are unaware that their own expressions may not necessarily be comparable to the controlled vocabulary of the system, or that perhaps they find it difficult to make such multiple transformations. Lancaster and Baker (ibid, p.209) cite several studies to highlight the extent of the problem:

> The initial failure experienced by a patron when trying to match a query term to a catalog subject heading would not be so important if the patron would then try other subject headings. A few highly motivated patrons will actually do so... The typical patron, however, consults only one subject heading. If that term has no matches, he or she either assumes the library had no materials on the subject or simply gives up.

Steinberg and Metz (cited in Markey, 1984) found that only 28% of the users surveyed were aware that subject searching required them to match an LCSH heading. Whilst Bates (1986) examined this is practice by comparing the degree of match between users' query terminology and LCSH, and found that only just over 20% initially matched or partially matched the controlled LCSH vocabulary.

Therefore, although participants may have had a fairly clear idea of their search topic, owing to an assignment specification, they displayed some difficulty in transforming this topic into an appropriate query. Commercial OPAC designers have done little to ease this task. As Markey asserted in 1985, but as pertinent today:

> In our exisiting online catalogues, patrons are saddled with the job of finding the right subject heading to represent their topic of interest (Markey, 1985, p.40)

In a View-based system, the system indexing language is *displayed* in a logically structured hierarchical arrangement and therefore the user is not attempting to match *blindly* his or her own vocabulary to that of the system. Also included is a cross-reference structure that directs users to the preferred term if their own term choice fails to match.

Bates (1996) found that users tended to begin searching with a broader rather than narrower term in order "to survey the territory, to size up the system's offerings". She relates this to research into reference desk interactions, which finds that users tend to ask for more general information than they are actually seeking in order to 'test the water' and provide the librarian with a search context. She draws upon anthropological and linguistic research that points to the existence of folk classifications which display similar structure across languages and culture, suggesting that the human mind may have a set of generic categories from which such classifications arose. She believes that classifications such as Dewey have such 'folk classification' origins and perhaps display some of their properties (hierarchies with few levels, 250-800 generic categories). Whether they do and whether there is such a thing as a universal classification may perhaps be overly deterministic but more importantly, the development of world-wide folk classifications in the first place and the numerous examples of classification we find in every day life (a supermarket layout, for example) suggests that humans may be naturally disposed to categories of information, organised from the general to the specific. A View-based system provides such a familiar context.

Despite participants' frequent term mismatches and lack of variety in their actual search concepts, they did display an awareness of semantic relationships such as hierarchy and equivalence when attempting to refine their search. The nature of subject terminology *progression* may be worthy of further investigation since it provides an insight into users' concepts and indicates their awareness of semantic relationships for broadening or narrowing a search. Examining the progression of subject search terms recognises that this choice is a sequential process and does not treat terms in isolation of each other. These classificatory relationships are utilised in our view-based searching OPAC and are central to its understanding and effective use.

The notion of *what constitutes a 'successful' subject search* is much more complex than for known-item searching. In section 3.3.1, known item searches were judged successful if the sought after item was retrieved directly following a single query formulation. In subject searching, the user does not have a specific item in mind to help fulfil his information requirement so in a sense there is no clear 'right' or 'wrong' answer to the posed query. As was discussed earlier in relation to search goals/purpose, the subject searching of the participants in this study was generally task-driven by specific assignment questions and not by a vague exploration of subject matter. A successful search outcome would probably retrieve items that in the eyes of the participant would go some way towards answering such questions, and they may not be able to determine whether the search outcome has been successful or not until they have fully consulted the item. This did seem to be the case, as

participants were keen to jot down class codes, paying little attention to subject indexing information, and finalise their selection at the library shelves.

Baker and Lancaster (1991) note that subject search success has been defined narrowly, in terms of an exact match with the indexing vocabulary, and more broadly such as whether users would be able to locate a relevant subject headings (perhaps via cross-references or alphabetically similar heading). However, they also point out that these studies did not observe real users. In addition to the degree of term match, the amount of effort expended is also a factor in search success; for example, whether the volume of information retrieved by the search strategy is manageable and if not, is it easily reducible? To what extent does the system facilitate this? The notion of manageability is to some extent subjective and relates to the user's search goals and persistence.

As a result of research into the statistical properties of information stores and access methods, the Resnikoff-Dolby Rule suggest 30:1 as the optimum ratio for human information processing (Bates, 1996). This rule was later investigated by Wiberleym, Daughery and Danowski (cited in Bates, 1996) to examine the number of hits inspected by users at the OPAC interface, finding a 'typical persistence' level of 30-35 hits viewed and a 100-200 threshold. Incidentally, the Horizon OPAC interface at Huddersfield upholds this by retrieving 30 references in the first instance for subject keyword searches, with the option to 'Retrieve More'. However, these first 30 hits appear in no particular order, unless the user opts to sort them alphabetically by subject heading or numerically by items per title; the system does not support relevance ranking and/or feedback. These sorting options do not then allow a user to modify a large result set with further criteria (only re-order the same number of items) for this the user must reformulate the query. A view-based OPAC allows users to refine a search persistently with a number of criteria (e.g. 'Date of Publication', 'Type of Material') and each to varying degrees.

Particularly pertinent in subject searching is the need for users to distinguish between a 'collection failure' and 'search failure' yet this is not made explicit by the current OPAC system. Users may be unsure as to whether their choice of search term failed to match the indexing language of the system (meaning that the library may in fact hold books on the particular topic only represented by differing terminology); or that the library holds nothing at all for the given topic. The situation remains ambiguous. In the Horizon OPAC (and many other current library systems), there is little accommodation of variability in search term choice. If users fail to match a query term in a subject keyword search, users are instructed that: *the following word(s) is (are) not indexed. Try other related word(s), or check the spelling*. The distinction between collection and search failure is perhaps made more transparent in a view-based system, since the user can see the system indexing vocabulary, and thus can view how a topic is represented and the number of (including zero) items pertaining to it in the library. In cases where the user directly searches the controlled vocabulary and fails to make a match then, as mentioned previously, the provision of cross-referencing tables directs them to the preferred term and its location in the subject tree. Here the user can again see whether or not the library holds items on the sought-after subject and related subjects, enabling search refinement as appropriate.

Determining the relevance of information items is also perhaps more complex for subject searching than for known-items. It was noted in the earlier discussion that by its nature, the relevance of known-items was generally determined prior to searching, typically by a lecturer. In subject searching, the onus is placed more firmly on the student, although usually with a specific goal in mind.

The majority of participants reported that they had found items by *browsing* the library shelves. The views expressed by participants seemed to suggest that they regarded the OPAC primarily as a tool for *finding* specific items and not for discovering and exploring information, although, as discussed, the users in the present study tended to have explicit goals which may not leave them as open to browsing. Shelf browsing for these participants was often instigated due to a failed OPAC search or if users perceived that the OPAC presented insufficient information to enable a relevance judgement. Marchionini (1995) suggests several reasons for browsing:

- To gain an overview of a physical or conceptual space
- To monitor a process (e.g. current awareness journal scanning)
- Requires a smaller cognitive than query formulation
- To clarify information problem (examine indexing vocabulary, develop concepts)
- To discover and learn (serendipity, cross disciplinary boundaries)
- Environment invites browsing

Hildreth (1995) characterises browsing as "a family of information seeking activities" and collates the typologies of browsing behaviour found in the literature. These typologies comprise three basic levels, distinguished by (a) the extent to which the user has a specific goal in mind and (b) the systematicity of tactics employed (Marchionini, 1995):

Category 1	Category 2	Category 3
undirected browsing	semi-directed browsing	directed browsing (Herner, 1970)
general browsing	general purposive browsing	specific browsing (Apted, 1971)
serendipity browsing	general purpose browsing	search browsing (Cove & Walsh, 1988)

From Hildreth (1995)

Category 1 browsing is "largely random, unstructured and undirected"; category 2 constitutes habitual browsing, for example of a particular publication for current awareness; whilst a category 3 browser "has a specific end in mind, but does not approach the catalog with a well-formulated search strategy" (Hildreth, 1995). If we examine the study findings in light of this typology then subject searching participants can be viewed as engaging in Category 3 directed/specific/search browsing at the OPAC interface: they displayed uncertainty in their query formulation and search strategy yet had a specific idea about the nature of the desired output. The other browsing behaviour types were not observed yet are perhaps more conducive to the physical environment of the library shelves, particularly in view of current OPAC design.

Current commercial OPAC interfaces are modelled on the traditional query-matching paradigm with its command-driven interfaces. Users are seen as having a clear and unchanging information need which they are able to formulate into a precise Boolean query to match and retrieve a single ideal set of documents represented by the search terms. Recent thinking, however, claims that users approach the system with uncertain information needs, which may develop and change as the user encounters information throughout the interaction (what Bates terms 'an evolving search'); the search path is not necessarily systematic but iterative and heuristic. Although, like Hert (1997), in this study it was found these participants' goals did not change considerably *during the course of the OPAC interaction*, it is recognised that any OPAC interaction is embedded within a larger information seeking process during which users may consult multiple different information resources that may lead them in alternative directions. This is perhaps more likely in the case of researchers, who have more academic freedom than the majority of undergraduate participants in this study whose searches are constrained by assignment requirements and deadlines thus precluding significant changes in information need. Looming deadlines seemed to be at the forefront of participants' minds, providing a structure and restraint in their information seeking behaviour. As discussed earlier, despite failing to support the finding that users displayed no substantial change in their information needs when considered solely in terms of the OPAC interaction, users did have problems translating their subject search needs into an explicit query. These subject searching participants can be viewed as category 3 directed/specific/search browsers yet the Horizon OPAC interface did little to support them in this activity

Organised environments facilitate and support browsing by displaying and relating similar items (Marchionini, 1995). For electronic environments, Marchionini lists numerous techniques to support browsing. Those incorporated into the view-based searching interface include conceptual displays, use of physical features (e.g. icons to indicate type of document), direct manipulation and query by selection/recognition via hierarchies. Like the majority of commercial OPACs, the Horizon system at Huddersfield includes none of the browsing support features mentioned by Marchionini. The Web version is slightly more amenable as it includes hypertext subject headings, enabling the user to find a record then cross-reference to similar titles. However, in the study with the library-based OPAC, participants seemed to pay little attention to such subject headings. The Horizon WebPac, however, highlights these subject headings by hypertext which might perhaps direct users' attention to these and encourage their use.

In this study, those participants engaged in subject searching (or directed/specific/search browsing), as opposed to known-item searching, would have perhaps benefited from having browsing support features incorporated into the OPAC. It may have helped them clarify their information needs by viewing and exploring the indexing vocabulary, and better facilitated the iterative search process that was evidenced by their query reformulation tactics than the current cumbersome 'one-shot' discrete query approach, which seems inappropriate under such circumstances. Hildreth (1995) amongst others (e.g. Bates 1984, Kuhlthau 1997) urges that OPAC designers:

> break out of the query-oriented, Boolean mind-set, we need to turn the conventional query-first-then-browse paradigm upside down. Searching by exploration recognition, and discovery in a well structured bibliographic space should be the primary search interface provided to information seekers...

5. A few thoughts on further research

Both quantitative and qualitative measures proved insightful in this pilot study of OPAC-user interaction. Quantitative measures gave an indication of user group characteristics, the relative frequencies of search types and the chosen indexes (access points). Qualitative data illuminated in context the sorts of problem experienced by participants and their thoughts on these. Users' understanding and development of subject searches was examined using a basic linguistic analysis of the query formulation and reformulation process in respect of search term choice, grammatical structure and the semantic relationships displayed.

Future research might incorporate different types of user. The participants in this study were predominantly undergraduates whose OPAC searching was driven and seemingly strongly influenced by assignment criteria. Other user groups such as researchers and lecturing staff may be using the OPAC, perhaps from their own desktops rather than the library building, with different types of search goals and information needs. This relates to the generalisability of the study.

Improved methods (i.e. audio-visual) for capturing the interview and OPAC-user interaction data would also be extremely valuable, allowing more detailed data analysis.

An updated product review on the current state of commercial OPAC interface design might also be useful.

The challenges of subject searches proved most interesting in this study. Possible future questions could include: What are the users' objectives/goals when subject searching? Do these goals change during the interaction and what motivates this change/no change? What are the sources of users' subject query terms? How do users' own terms relate to those used in controlled vocabularies? What are the common problems and why do these occur? How do users respond to the system feedback, i.e. how do they develop their search semantically, morphologically and syntactically and what does this tell us about their understanding of the semantic relationships that are central to query refinement? How can we address the nature of subject searching and the problems users experience in OPAC interface design?

References

Baker, S. L. and F. W. Lancaster (1991) *The measurement and evaluation of library services*, 2nd ed. Arlington, Virginia: Information Resources Press.

Bates, M. (1986) Subject access in online catalogs: a design model. *Journal of the American Society for Information Science* 37(6), 357-376.

Bates, M. (1996) *Indexing and access for digital libraries and the Internet: human, database, and domain factors* [Online] Available at: < http://dlis.gseis.ucla.edu /research/mjbates2.html> [Accessed 29 September 1999]

Hert, C. A. (1997) *Understanding information retrieval interactions: theoretical and practical implications*. London: Ablex Publishing Corporation.

Hildreth, C. R. (1995) Conclusion: outline of the next generation of online catalogs. In: *Online catlog design models: are we moving in the right direction? A report submitted to the Council on Library Resources August 1995.* [Online] Available at: <http://www.ou.edu/faculty/H/Charles.R.Hildreth/clr-five.html> [Accessed 15 February 2000]

Kuhlthau, C. C. (1999) Accommodating the user's information search process: challenges for information retrieval designers. *Bulletin of the American Society for Information Science* 25(3), 12-16.

Large, A. and Beheshti, J. (1997) OPACs: a research review. *Library and Information Science Research* 19(2), 111-113.

Marchionini, G. (1995) *Information seeking in electronic environments*. Cambridge: CUP.

Markey, K. (1985) Subject searching experiences and needs of online catalog users : implications for library classification. *Library Resources and Technical Services,* 29(1), 34-51.

Markey, K. (1989) Subject searching strategies for online catalogues through the Dewey Decimal Classification. *In*: Hildreth, C R (ed.) *The online catalogue; developments and directions.* London: The Library Association, pp.61-83.

Mills, J. (1997) Introductory address. *Knowledge organization for information retrieval, proceedings of the sixth international study conference on classification research held at University College London 16-18 June 1997.* The Hague, Netherlands: FID, pp.1-11.

O'Brien, A. (1994) Online catalogs:enhancements and developments. *In:* M Williams (ed.) *Annual Review of Information Science and Technology*, pp.219-24

Taylor, R. S. (1968) Question-negotiation and information seeking in libraries. *College and Research Libraries* 29, 178-189.

Walker, S. (1989) The OKAPI online catalogue research projects. *In*: Hildreth, C R (ed.) *The online catalogue; developments and directions.* London: The Library Association, pp.84-106.

Walker, S. (1991) Subject access in online catalogues. *In: OPACs and the user. Proceedings of the third Anglo-Nordic Seminar 8-11 April 1990. British Library R & D Report 6040.* [London]: The British Library Board, pp. 23-32.

Appendix 2

Library Online Public Access Catalogue: System Usability Scale

Please evaluate the usability of the system *in relation to the four tasks*.

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

11. Please add any comments about the library system

_____

_____

_____

_____

Appendix 3 Usability Study SPSS Data

	system	freq	complex	ease	support	function	consist
1	1	3	4	4	4	3	2
2	1	4	2	3	2	3	3
3	1	3	3	4	5	4	5
4	1	3	3	3	5	2	4
5	1	3	3	4	5	4	4
6	1	4	4	3	4	5	4
7	1	4	4	3	4	4	4
8	1	5	5	5	5	5	5
9	1	4	5	4	5	4	5
10	1	4	4	4	2	3	4
11	1	2	2	2	2	2	3
12	1	3	4	4	5	3	4
13	1	4	3	3	4	4	2
14	1	4	4	3	4	5	2
15	1	4	4	4	5	3	4
16	1	3	4	3	5	3	4
17	1	4	3	2	3	4	4
18	0	5	5	5	5	3	5
19	0	3	1	3	5	3	2
20	0	4	4	5	5	4	5
21	0	4	5	5	5	4	5
22	0	2	4	3	3	3	3
23	0	4	5	3	5	3	3
24	0	3	2	2	2	5	3
25	0	3	2	2	3	3	2
26	0	4	4	3	5	2	4
27	0	2	5	3	5	4	2
28	0	3	2	2	2	1	3
29	0	4	4	5	4	4	5
30	0	5	4	5	5	5	4
31	0	3	4	4	5	2	2
32	0	4	3	2	4	3	2
33	0	3	4	4	5	5	4
34	0	4	5	5	5	4	5

Appendix 3 Usability Study SPSS Data

	learn	cumber	confid	starting	eou	Icon	fi
1	5	4	3	2	4	4	3
2	2	3	3	5	3	4	3
3	4	4	4	5	4	5	5
4	3	1	3	2	2	3	3
5	4	4	4	4	4	4	4
6	5	2	3	2	3	4	5
7	3	4	3	4	4	4	4
8	5	5	4	5	5	5	5
9	5	5	4	4	5	5	5
10	4	3	3	2	4	3	4
11	3	3	3	3	2	3	3
12	4	4	4	2	4	3	4
13	5	4	3	4	3	5	3
14	5	4	4	2	4	4	4
15	5	4	4	5	4	5	4
16	3	4	3	3	4	3	4
17	1	2	3	2	2	2	4
18	5	3	5	5	4	5	4
19	1	2	3	2	2	2	3
20	5	3	4	5	4	5	5
21	4	5	5	5	5	5	5
22	4	3	3	3	3	4	3
23	5	4	4	5	4	5	3
24	4	3	3	2	2	3	4
25	3	2	3	2	2	3	3
26	3	4	3	4	4	4	3
27	3	2	4	5	3	4	3
28	2	2	2	1	2	2	2
29	4	3	5	5	4	5	5
30	4	3	5	5	4	5	5
31	4	5	4	5	4	5	2
32	4	3	3	4	3	4	3
33	5	1	5	5	3	5	5
34	5	5	4	5	5	5	5

Appendix 3 Usability Study SPSS Data

	sc	us	sum	t1time	t2time	t3time	t4time
1	4	3	34	.	466	284	184
2	3	4	30	280	258	215	100
3	5	3	41	410	119	219	92
4	4	3	29	351	346	296	132
5	5	3	39	164	282	231	149
6	4	4	36	219	436	450	209
7	4	4	37	150	188	188	108
8	5	5	49	285	325	200	167
9	5	4	45	396	0	828	250
10	3	4	33	607	0	205	219
11	3	2	25	589	617	224	439
12	5	3	37	700	540	141	433
13	4	4	36	408	82	294	124
14	4	4	37	503	91	131	75
15	5	4	42	305	89	164	124
16	4	3	35	717	555	95	155
17	3	4	28	331	302	102	127
18	5	5	46	453	129	312	181
19	4	3	25	319	203	243	264
20	5	4	44	243	231	256	286
21	5	4	47	247	278	44	289
22	3	2	31	844	191	77	93
23	5	4	41	1062	253	112	1357
24	3	3	29	1052	357	425	681
25	3	3	25	873	397	587	430
26	4	4	36	646	285	505	330
27	5	2	35	820	114	179	309
28	2	3	20	265	52	222	250
29	5	4	43	723	327	237	315
30	5	5	45	265	258	240	387
31	5	3	38	472	454	282	628
32	4	4	32	417	505	318	640
33	5	3	41	566	441	403	574
34	5	4	47	304	96	139	179

	overall
1	311
2	213
3	210
4	281
5	207
6	329
7	159
8	244
9	369
10	258
11	467
12	454
13	227
14	200
15	171
16	381
17	216
18	269
19	257
20	254
21	215
22	301
23	696
24	629
25	572
26	442
27	356
28	197
29	401
30	288
31	459
32	470
33	496
34	180

Appendix 4

Participants' Qualitative Comments about the Library Systems

Participant	System	Comment
1	VBS	I think I need more categories to narrow my search. I got confused in the end while reading all the search results. The system may work for a specific subject but no the whole catalogue.
2	VBS	
3	VBS	It would be interesting to see the final program.
4	VBS	More info needed in the details box. Categories for refinement sometimes not helpful – needed to practice to be effective. Good idea. (I failed to understand what was meant by literature review!)
5	VBS	Once we understood how to use it, it was quite easy to use.
6	VBS	Any problems could be overcome with more practice. This will be really useful for students. When you do a search – more explicit instruction would be better, i.e. Type in a word NOW HIGHLIGHT ITEM IN POP UP LIST.
7	VBS	I used to be a librarian so am used to the Dewey system and have some understanding of the structures behind the system so I adapted fairly well to this new system.
8	VBS	This system would be extremely helpful for people who had no idea what particular books they were searching for. I have found a few titles which I had not heard of which would be particularly useful for me.
9	VBS	Some subject headings were unclear. An overview of each resource would be helpful.
10	VBS	I think once you familiarise yourself with the system, you feel comfortable. There is some need for tutorial, though a short one for potential searchers to use the system to its potential.
11	VBS	Lots of potential
12	VBS	The system gives you no help - if you get stuck it's up to you to backtrack. Perhaps an eventual addition would be a pop-up menu, offering to narrow or broaden your search options?
13	VBS	Main problem came with the top LH box, not realising that something had to be highlighted before you could type in the subject topic and then pressing 'go'. I hadn't realised you could type in the box! (I must have not been concentrating when that was explained at the beginning!). With a bit more practice, I would really enjoy using it and find it very efficient. Enjoyed it! Thank you!
14	VBS	I an reckoning that this system is built on an SQL queried database. As I am competent in SQL it was easy to

		translate these skills onto this system. Perhaps this helped when calculating what the headings and subheadings I needed were. This system made it easier to find information than the current system but some headings were to specific, i.e. no heading for pure maths only subsystems for Algebra etc. So you need a clear idea of what you are looking for.
15	VBS	The above evaluation is in the light of having needed a practice go in which I got a bit lost. Once I had had a go and learned where I'd gone wrong, I was confident in using the system and found it easy to find my way around.
16	VBS	Would need a good set of instructions with step-by-step guidance and several practice examples.
17	VBS	Navigation was difficult at start. Why the two colours – beige and green?
18	IPAC	If there was a short description of books when selected, this would enable users to gain a better insight as to what the text is about, rather than just going by the title.
19	IPAC	It would be nice if it offered "similar" results or "guessed" what you were looking for – suggestions based on topic/subject ("see also"). A grouping by subject?
20	IPAC	It was a good experience. System is very simple to use. I think spell check algorithm should be there to save users time.
21	IPAC	It is sometimes impossible to get any articles it the search criteria is specialised. Maybe it should be re-designed to include a search engine component like Google where it will show articles related to the search components if they are specialised.
22	IPAC	The sample questions asked were of a literary nature – not problems in finding answers if you know something of the subject. Someone new to this or any subject would find the system difficult to use.
23	IPAC	- Difficult to locate items that have more than 2 parts to the search, e.g. literature review + Higher Ed. Inst. + West Yorks. - Searching by "subject alphabetical" is very useful to pinpoint particular subject areas. - Would be helpful if the system could make more lateral suggestions, particularly when searching unfamiliar subject areas.
24	IPAC	- All these searches required at least two levels of searching and then evaluating the items found. This proved rather time-consuming and disappointing at times. - The spec(?) of the search it "type of search dependent . Quite often one had to settle for the closet items. There is some measure of uncertainty.
25	IPAC	1) Inconsistency- the search brought up topics titled "English poets- 19th century" however if keying into one search filed the same criteria the computer

		could not find any search results. 2) Knowledge around the subject matter being searched for is required. 3) Had difficulty understanding how the AND/OR options work.
26	IPAC	When you know what you are looking for, e.g. names and dates, the system will retrieve the information. It is not very useful when trying to do a search using only generalised information. Not a supportive system.
27	IPAC	Better se? so that here is more fluency in the work. Must be possible to only use the keyboard. Try building a subject index with root system. And definitely add abstracts to the different works listed.
28	IPAC	This is not an intuitive system – why else would they employ and "Ask a Librarian" tag. Adjectives – clumsy, disjointed, sparse. It does not seem to assist your search with suggestion in an Amazon style.
29	IPAC	
30	IPAC	Easy to use and helpful for easy browsing.
31	IPAC	More keywords must be included in the search function, As an MSc. Engineering student, when I search the catalogue I found out many keywords missing. It took me some time to find certain books. The keywords should be more related to the subject than author or title. Thanks for letting me do the study.
32	IPAC	A spell checker should be provided. Advanced/expert searches should be improved.
33	IPAC	System is really flexible = to search library catalogue. All the functions are easy to access.
34	IPAC	It would be more helpful if a small description was included with the books instead of just the author and title. The title can be a bit vague and this would help.

Appendix 5: IPAC Search Terms

Task 2: Find 3 biographical works about 3 different classical composers

Search Terms	Index	Number of Participants Using Terms	Number of hits
Music	Subject Keyword	1	9948
music, composers	Subject Keyword	1	378
biographical works composers	General Keyword	1	0
biographical composers	General Keyword	1	17
biographical works composers	General Keyword	2	0
biography classical composer	General Keyword	2	0
biography composer	General Keyword	2	24
biography classical composers	General Keyword	5	1
boigraphies of classical composers	General Keyword	1	0
classical biographies	General Keyword	1	1
classical composers	General Keyword	9	6
classical music biographies	General Keyword	1	0
biographies	General Keyword	1	1268
mozart	General Keyword	2	881
Music- biographies	Subject Keyword	4	2
Mozart , Wolfgang Amadeus , 1756-1791	Subject Keyword	2	104
Musicians - Biographies	Subject Keyword	1	3
musician biographies	Title Keyword	1	0
Composders	General Keyword	1	0
Composers	General Keyword	2	1088
Composers - Biography	General Keyword	2	188
Composers - Individual Biographies	Subject Keyword	2	2
Classical Composers - Individual Biographies	Subject Keyword	1	0
Classical Composers	Subject Keyword	1	0
Classical Composer	Subject Keyword	3	0
Classical Composer	General Keyword	1	0
Beethoven	Subject Keyword	1	101
Schubert	General Keyword	1	457
Bach	General Keyword	1	985
Bach biography	General Keyword	2	7
Beethoven biography	General Keyword	2	10
Classical composer	General Keyword	1	0
Composers	Subject Keyword	1	781
Biography Chopin	General Keyword	1	4
types of classical composers	General Keyword	1	0
biographical work on classical composers	General Keyword	1	0
Mozart's biographical works	Author Keyword	1	0
biographical works	Author Keyword	1	0
Mozart	Author Keyword	1	741
classical comp9osing	Author Keyword	1	0
classical composing	Author Keyword	1	0
Beethoven	Author Keyword	1	538

Schubert , Franz , 1797-1828	Author Keyword	1	375
biographical works classical composers	General Keyword	1	0
biographical classical composers	General Keyword	1	0
biographical works classical	General Keyword	1	0
biographical works classic	General Keyword	1	0
biographical works	General Keyword	1	4
classical composers biographical	General Keyword	1	0
classical composers biographical works	General Keyword	2	0
classical composers works	General Keyword	1	0
classical composers biographic	General Keyword	1	0
biographical classical composers	Title Keyword	1	1
biographical works about 3 classical composers	General Keyword	1	0
biographical works of classical composers	General Keyword	1	0
works of classical composers	General Keyword	1	0
biographic work about classical composers	General Keyword	1	0
work classical composers	General Keyword	1	0
graphical work of classical composers	General Keyword	1	0
3 classical composers	General Keyword	1	1
biographical works about music composers	General Keyword	1	0
biographical works about music players	General Keyword	1	0
biographical works about piano	General Keyword	1	0
composer biography new grove	General Keyword	1	11

Task 3: Find 3 biographical works about 3 different 18th century poets

Search Terms	Index	Number of Participants Using Terms	Number of hits
English poetry	Subject Keyword	1	1512
18th century poets	General Keyword	5	9
18th century poet	General Keyword	1	1
18th century poetry	General Keyword	2	56
English poetry - 18th century	Subject Keyword	3	49
18th century english poets	General Keyword	2	9
poets	General Keyword	1	335
poets 18th century	Title Keyword	1	0
poets	Title Keyword	1	183
18th century	General Keyword	1	938
German poetry - 18th century	Subject Keyword	1	5
english poertry-eighteenth century	Subject Keyword	1	0
english poetry-eighteenth century	Subject Keyword	1	0
English poetry	Title Keyword	4	121
english poetry 18th century	Title Keyword	3	1
English Poetry (1702-1745) - Anthologies	Subject Keyword	1	3
18th century english poets	Subject Keyword	1	3
ancient mariner	Subject Keyword	1	1
Coleridge , Samuel Taylor , 1772-1834 , Rime of the ancient mariner	Subject Keyword	1	1
Coleridge , The rape of the lock	Subject Keyword	1	0
The rape of the lock	Subject Keyword	1	2
Milton John	Subject Keyword	1	123
Keats	General Keyword	1	80
Romanticism - English Poetry (1800-1837)	Subject Term Selection	1	3
18th cenrury poetry	Subject Keyword	1	0
18th century poetry	Subject Keyword	1	55
Chatterton , Thomas , 1752-1770	Author Term Selection	1	

18th century	Title Keyword	1	41
Byron	Author Alphabetical	1	88
Byron , George Gordon Byron , Baron , 1788-1824	Author Term Selection	1	26
Wordsworth	Author Alphabetical	1	46
Wordsworth , William , 1770-1850	Author Term Selection	1	33
Coleridge	Author Alphabetical	1	57
Coleridge , Samuel Taylor , 1772-1834	Author Term Selection	1	37
english poetry in 18th century	General Keyword	1	10
english poetry	General Keyword	2	1608
Wordsworth	Author Keyword	1	46
English poetry - History and criticism	Subject Term Selection	1	239
english poetry 18 century	General Keyword	1	2
english poetry 18th century	General Keyword	1	52
Poetry	Subject Keyword	1	1810
english poetry eighteenth century poet	Title Keyword	1	0
eighteenth century poet	Title Keyword	1	0
eighteenth century poets	Title Keyword	1	2
works of 18th century poets	General Keyword	1	0
works of english poetry	General Keyword	1	57
3 works of english poetry	General Keyword	1	2
3 works of english poets	General Keyword	1	1

Task 4: Find 3 items discussing the work of English female novelists in the 19th century

Search Terms	Index	Number of Participants Using Terms	Number of hits
english novels	Subject Keyword	1	7
19th century female novelist	General Keyword	1	0
19th century novels	General Keyword	2	23
19th century female writers	General Keyword	1	0
19th century witers	General Keyword	1	0
19th century novelists	General Keyword	3	34
Novelists, English - 19th century - Biography	Subject Term Selection	5	18
19th century female novelists	General Keyword	4	0
19th century female novelists	Subject Keyword	1	0
19th century novelist	Subject Keyword	1	0
19th century	General Keyword	2	2332
19th century victorian novels	General Keyword	1	6
Women authors, English - 19th century	Subject Term Selection	1	40
English female novelists	Subject Keyword	1	0
English novelists	Subject Keyword	1	49
Eliot , George , 1819-1880	Subject Term Selection	1	58
Victorian novels - English fiction (1837-1900) - Texts	Subject Term Selection	1	5
Bronte , Emily , 1818-1848 , Wuthering heights	Subject Term Selection	1	10
English fiction (1837-1900)	Subject Term Selection	2	380
Women and literature - England - History - 19th century	Subject Term Selection	2	22

Shelley , Mary Wollstonecraft , 1797-1851 - Criticism and interpretation	Subject Term Selection	1	2
Women authors, English - 19th century - Biography	Subject Term Selection	1	3
Eliot , George , 1819-1880 - Criticism and interpretation	Subject Term Selection	1	10
English fiction - 19th century	Subject Keyword	1	356
English fiction - 19th century - History and criticism	Subject Term Selection	1	134
English fiction - 19th century	Subject Alphabetical	1	226
English novelists	Subject Alphabetical	1	0
English novelists	General Keyword	5	70
Delafield , E. M. , 1890-1943	Author Term Selection	1	1
simone de beauvoir	Author Keyword	1	25
Women - Sociology	Subject Term Selection	1	102
Bronte	Author Keyword	1	41
female novelists	Subject Keyword	1	0
Novelists	Subject Keyword	2	70
Bronte	General Keyword	1	105
Pride and Prejudice	General Keyword	1	21
Shelley Mary	Subject Keyword	1	22
english female novelist 19th century	Subject Keyword	1	0
english 19th century	Subject Keyword	1	727
Nesbitt , Cathleen	Author Term Selection	1	1
Walker , Zena	Author Term Selection	1	2
Plowright , Joan	Author Term Selection	1	1
english 19th century literature	General Keyword	1	259
english 19th century	General Keyword	1	774
english 19th century literature female	General Keyword	1	3
!9th century novelists	Title Keyword	1	0
Discussion	Subject Keyword	1	21
19th Century novelists	Title Keyword	1	0

19 novelists	Subject Keyword	1	0
19th novelists	Subject Keyword	1	27
Ritchie , Anne Thackeray , 1837-1919 – Biography	Subject Term Selection	1	1
female Novelists, English - 19th century – Biography	General Keyword	1	0
Gerin , Winifred	Author Term Selection	1	6
female novelists in 19th century	Title Keyword	1	0
female novelists	Title Keyword	1	0
novelists	Title Keyword	1	34
English fiction - Women authors - History and criticism	Subject Term Selection	1	46
english female novelists	General Keyword	5	0
english female novelist	General Keyword	2	0
female novelist	General Keyword	1	0
novelist	General Keyword	1	25
female novelists	General Keyword	5	0
enlgish novelists	General Keyword	1	0
Women authors, English - History and criticism	Subject Term Selection	1	116
Austen , Jane , 1775-1817	Subject Term Selection	1	49
Gaskell , Elizabeth , 1810-1865 - Criticism and interpretation	Subject Term Selection	1	10
Waugh , Evelyn , 1903-1966	Subject Term Selection	1	14
Hardy , Florence Emily	Subject Term Selection	1	1
Novelists, English - 19th century - Wives	Subject Term Selection	1	1
Atwood , Margaret , 1939- - Criticism and interpretation	Subject Term Selection	1	1
Shelley , Mary Wollstonecraft , 1797-1851	Subject Term Selection	1	20
19th century english female novelists	General Keyword	1	0
Novelists	General Keyword	1	109
novel female	General Keyword	1	3
novel female english	General Keyword	1	1
19th century english female	General Keyword	1	3

ninteenth century english female novelists	General Keyword	1	0
ninteenth century english female novelists	Title Keyword	1	0
english novel	General Keyword	1	200
english novelists ninteenth century	General Keyword	1	0
english novelists 19th century	General Keyword	1	27
english novelists women 19th century	General Keyword	1	4
englihhh	General Keyword	1	0
work of english female novelists	General Keyword	1	0
work of female novelists	General Keyword	1	0
English women novelists	General Keyword	1	9
work of english women novelists	General Keyword	1	0
works of novelists	General Keyword	1	1
works of english novelists	General Keyword	1	0
work of english novelists	General Keyword	1	0
work english novelists	General Keyword	1	4
work of novelists	General Keyword	1	0
Female 19th century	General Keyword	1	7
19th century female	General Keyword	1	7
female writing	General Keyword	1	2
female nineteenth	General Keyword	1	2
women nineteenth	General Keyword	1	38