



# University of HUDDERSFIELD

## University of Huddersfield Repository

Binzabiah, Reyad and Wade, Steve

An Approach to Developing an Ontology that Represents Knowledge Embedded in Filmed Materials

### Original Citation

Binzabiah, Reyad and Wade, Steve (2012) An Approach to Developing an Ontology that Represents Knowledge Embedded in Filmed Materials. In: Proceedings of The Queen's Diamond Jubilee Computing and Engineering Annual Researchers' Conference 2012: CEARC'12. University of Huddersfield, Huddersfield, pp. 91-97. ISBN 978-1-86218-106-9

This version is available at <http://eprints.hud.ac.uk/id/eprint/13457/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: [E.mailbox@hud.ac.uk](mailto:E.mailbox@hud.ac.uk).

<http://eprints.hud.ac.uk/>

# An Approach to Developing an Ontology that Represents Knowledge Embedded in Filmed Materials.

R. Binzabiah , Dr. S. Wade  
Informatics Department, School of Computing & Engineering  
University of Huddersfield  
[r.binzabiah@huddersfield.ac.uk](mailto:r.binzabiah@huddersfield.ac.uk) ; [s.j.wade@hud.ac.uk](mailto:s.j.wade@hud.ac.uk)

## ABSTRACT

*This paper introduces the reader to the approach we are taking to develop an ontology that could be used to represent the knowledge inherent in filmed materials. Such an ontology could be used as the semantic basis for multimedia retrieval systems. The proposed approach to ontology development is informed by the earlier work of researchers into folksonomy development and facet analysis. A brief survey of this earlier work is presented before our approach is described.*

**Keywords** Ontology , Folksonomy , Knowledge representation , Taxonomy , Information retrieval

## 1 INTRODUCTION

Thomas H. Davenport and Laurence Prusak in their pioneering book (Davenport and Prusak, 1997) adduce citation of a senior officer who was astonished by what he found, late in his career, in Tolstoy's famous novel War and Peace. He was surprised when he discovered how the description of the Napoleonic wars was vivid and full of details. The officer, after long experience in military corps and studying with military academics revealed that he had discovered a rich knowledge in the book which deepened his understanding of Napoleonic battles in a way that would have been impossible in the classroom. There is no doubt that, beside the stories, the cinema, TV, and all other moving visual materials play a significant role in our contemporary culture. On the one hand, they form our consciousness of the present world, and on the other hand, they can be considered as historical documents containing information that can shape our knowledge about the past. The question here is whether there is any knowledge that can be captured from filmed materials such as movies, video clips, documentary films, TV programs or even shots that were taken for news or film making purposes... etc. If the answer is "yes", this suggests a research agenda focused on ways in which this embedded knowledge should be captured, represented and made available to be used by those who need it. In the meantime, what is the tool that can be used to control this knowledge? One approach to achieve this goal is the tool called Ontology.

## 2 Literature Review of the Research

The background of this research project is comprised of several main aspects. The first of those aspects is knowledge management. The key concept of knowledge management is the sharing of knowledge between people to generate new knowledge (Davenport and Prusak, 1997), which could not be achieved without techniques for capturing, organizing, storing and retrieving knowledge held in various materials. Knowledge management techniques have also been applied to "intangible Assets" such as intellectual capital (Davenport and Grover, 2001). Here the emphasis is on developing a culture of learning which is the basis for innovation (Nonaka and Takeuchi, 1995).

The second aspect is the role of story and storytelling in knowledge management.

There is another phase of this research project related to knowledge representation (KR). If there is a meaning (or subject) contained in filmed materials that could form an interest for somebody, then this knowledge should be represented. Knowledge representation is a concept called to the "field of study concerned with using formal symbols to represent a collection of propositions believed by some putative agent" ((Brachman and Levesque, 2004):p4). It is "...most fundamentally a surrogate, a substitute for the thing itself, used to enable an entity to determine consequences by thinking rather than acting, i.e., by reasoning about the world rather than taking action in it."((Davis et al., 1993):p17).

“...ontologies aim to capture consensual knowledge in a generic way, and that they may be reused and share across software applications and by groups of people.” ((Gomez-Perez et al., 2004):p8) Therefore, the ontology will be the most significant phase of this research project.

The main reason behind choosing the ontology as a representation tool of knowledge embedded in filmed materials is to develop an information retrieval tool for these materials found in various types, since ontology can achieve this goal beside many others. Ren and Bracewell mentioned many efforts to tackle the subject of Multimedia Information retrieval (MIR), video retrieval particularly (Ren and Bracewell, 2009). In spite of this, there is increasing need for an accurate retrieval system which can match the growing amount of materials available either on-line or in the archives. Some of the most recent approaches tackle content-based, or semantic, retrieval. The main methods to perform semantic information retrieval (IR) are ontologies (Ren and Bracewell, 2009). These topics what will form one of multi aspects of this research project.

It is argued that there is a strong relationship between ontology and classification, to the extent that some scholars claim that they are the same (Dagobert, 1999). In fact, the concepts at any ontology should be constructed based on hierarchy taxonomy. This method has been followed in the field of Librarianship since the ancient world. Both, the ontologies and classification schemes provide coherent construction to organize the resources in effective manner (Choi, 2008a). Hence, this project will adapt some of the classification basics to construct taxonomy appropriate the filmed materials, particularly, when we consider that this project attempts to deal with the subjects contained in these materials. Therefore, taxonomy and classification will compose an important part in this research background.

Last part in this panorama of topics related to this project is the cinema and all what relates to motion picture creation process.

### 3 Related Works

This work project has relations with many other works. It is inspired by several ideas and projects which have tried to improve the process of retrieving filmed materials by the knowledge embedded in them.

#### I. Dian Tjondronegoro and Amanda Spink's study (Tjondronegoro and Spink, 2008)

This study has showed to any extent that the shortage in the quality of retrieval systems effected the search and retrieve multimedia materials, videos in particular, in spite of the massive increase in these materials. In addition to, that, the recent studies concentrate on the process of automatic annotation using the features such as (color, texture, shape...est.), the content-based approach, which needs machine learning and interpretation by using a modeled domain knowledge (e.g. Ontology) to be an intermediate to fill the gap between the users textual queries as a semantic meaning, and video data by observing visual features. For example, these features will not be efficient in searching certain subjects (e.g. emotions, relationships, or even Iraq War) (Tjondronegoro and Spink, 2008). Hence, the text is still the main method used to search huge amount of multimedia materials, but the problem as the study describes “Semantic concepts are still very limited to keywords without a structured vocabulary...” (Tjondronegoro and Spink, 2008) P347, and finally. the study recommended three things: (Tjondronegoro and Spink, 2008) P353.

- 1- “For semantic-based descriptions, there should be a standardized multimedia descriptions or ontology to unify the indexes”
- 2- “A well-developed folksonomy (community-based taxonomy) can act as a shared vocabulary that is both originated-by and familiar-to its primary users to make information increasingly easier to search, discover, and navigate over time”
- 3- “topic-based clustering of related videos will be very useful to assist users in navigating the multimedia collections”

#### II. Antonella Carbonaro work

Carbonaro (2008) presented an ontological approach providing a framework for semantic video retrieval and browsing which could assists users to access the content. Depending on the ontology, the systems dealt with the key entities and derived the relations. “An ontology-based

knowledge representation could be used for content analysis and concept recognition, for reasoning processes and for enabling user-friendly and intelligent multimedia content search and retrieval.”(Carbonaro, 2008) P203 based on word sense representation.

Analyzing the content by annotating it manually or automatically, enables to index the information retained in video. That's can be achieved by annotating related texts, subtitles, references, explanations, comments, closed captions or audio transcripts to transform these descriptions into instances in the ontology. This ontology-based system will be used in two faces, in content analysis, and in object recognition.

Also, this work tried to exploit the keyword approach to enhance user browsing and retrieving to introduce a collaborative approach can be beneficial for retrieving tasks and semantic coverage.

After several experiments to test the effect of the system “The ontology-based retrieving framework offers a valuable multimedia search functionality.” (Carbonaro, 2008) P211.

### III. The project of Electricité de France (EDF Recherche et Développement)

This Experimental project Designed to be implemented in EDF Company in France(Passant, 2007). The aim was to establish an internal weblog and group wikis platform to enable professional employees at the company to share their knowledge globally.

To index the contents of these weblogs, free-tagging method was introduced, which leads to appearance of Folksonomy, the collaborative technique widely used in social networks, to enrich the information retrieval capabilities. After one and a half year, with about 8000 posts and 80 bloggers and approximately 600 readers, some problems start to arise.

Therefore, Passant showed: “in order to offer better solutions to solve these variations and ambiguity problems - that are actually the reasons why controlled vocabularies as taxonomies have been introduced - and also offer a richer blog search experience to our users by suggesting related posts, we introduced a practice that kept the simplicity of free-tagging and added a formal layer on the top of the folksonomy, using semantic web technologies.” (Passant, 2007) The idea was to implement domain ontologies on the top of the folksonomy by linking each tag with the equivalent one or more concepts or instances in the ontology. Hence, in the opposite way, any resource could be linked with one or more tags. In meantime, users could add new tags or even link any existing concept in the ontology under the observation of platform administration.

## IV. Rubén Prieto-Díaz's Faceted approach to building ontologies

As long as the ontology is the centre of this research project, and since its quality depends on consistent building methodology, the faceted approach will be one of the most considered choices to construct the taxonomy, particularly when this ontology will tackle a very sophisticated area such as subjects of filmed materials. Thus, the approach proposed by Rubén Prieto-Díaz(Prieto-Díaz, 2003) is worth to take into account. This work describes using domain analyze method for building basis of any ontology, and borrows the library science approaches in identifying and categorizing concepts, and how the faceted method, which related to library science, incorporated into domain analysis method.

## V. Work of Christopher A. Welty \*, Jessica Jenkins (Formal ontology of subjects)

This work (Welty and Jenkins, 1999) is an attempt to apply an ontology of subjects devoted to a large scale digital library card catalog system. The important of this project is due to the focus on the retrieving system based on ontology of subjects of a digital library collection so that enables retrieve the materials based on its subject contents. The centre of this project was Model Editions Partnership (MEP)<sup>1</sup>. The goal of this project was to meet the sophisticated requirement of scholars with vague

---

1- maintainers of a large repository of fully marked-up electronic versions of historical documents from the US Civil War period. The Model Editions Partnership is a consortium of seven historical editions which has joined forces with leaders of the Text Encoding Initiative and the Center for Electronic Text in the Humanities. The participants have developed a "prospectus" setting forth editorial guidelines for publishing historical documents in electronic form. There are also seven miniature editions for those studying American history. [<http://www.teic.org/Activities/Projects/mo01.xml>].

The partnership brought together editors from seven on-going editorial projects and leaders from the Center for Electronic Texts in the Humanities (CETH) and the Text Encoding Initiative (TEI). Major funding for the three-year project, which began July 1, 1995, is provided by the National Historical Publications and Records Commission (NHPRC) with additional support from the University of South Carolina, Rutgers University and the University of Illinois at Chicago.[<http://www.dlib.org/dlib/november95/11chesnutt.html>]

notions about the object or recollection of search. These requirements may include queries such as: (works in certain genre in any kind of format, reviewed by some person who is a member of a certain organization who is interested in certain subject).

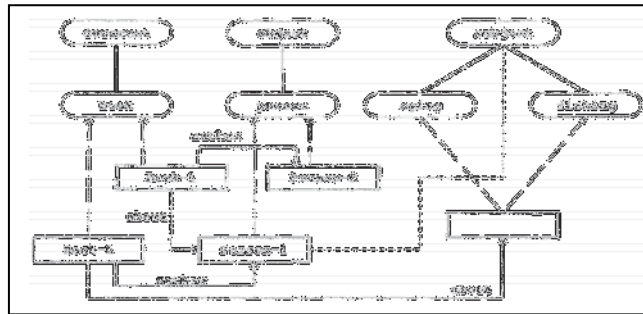


Figure 1 The structure of MEP library catalogue ontology. (Welty and Jenkins, 1999) P172

### VI. Vanda Broughton's study on faceted classification

What distinguishes this work (Broughton, 2006) is the answering of why should the faceted analysis and classification be in use when developing information retrieval systems. It is compared between various subject access tools (classification schemes, subject headings list, thesaurus, concept map and ontology) and detects the impact of facet analysis on them. In addition, it concentrated on commercial web sites to investigate the usage of faceted analysis technique of structuring information. Furthermore, this study focused on evaluating the technique of building a vocabulary and conceptual tool based on faceted analysis.

## 4 Research Context

The context of this research project will be conducted by a number of aspects. Firstly, building a lightweight ontology, this means avoiding use of a high degree of axiomatization richness in building the ontology.

“Lightweight ontologies usually are taxonomies, which consist of a set of concepts (i.e., terms, or atomictypes) and hierarchical relationships among the concepts.” (Zhu and Madnick, 2006):p1)

Secondly, this ontology will be restricted in filmed materials. This involve all the cinema, or video, movies and documents, in addition to any others similar items such as video clips and animated motion picture, either by computer or traditional ways such as cartoons, 2D or 3D.

Thirdly, the ontology will be text-base, deal with the subjects of these materials. This means that it will not contain the features of the videos such as color, shape, texture; length of the clips and scene breaks....etc. which will be described in this ontology as far as the need to. That is mainly because it will meet the needs for the subjective, mental needs, not the technical professional needs. Furthermore, it will describe the subjects of the resources. That is due to the fact that one of the characteristics of multimedia ontologies particularly, to describe the content. When it comes to the contents, there is a requirement to deal with subjects to regardless of the format of resources. (Seremeti and Kameas, 2007)

Next, this ontology will be designed to be developed by tagging system. This is mean the taxonomy will be a hierarchy of free tags, not subjects, in the sense of a thesaurus of tags not a classification scheme of subjects.

On the other hand, the taxonomy will take into account the features of folksonomy when designing it to be similar to folksonomy than the taxonomy.

Taxonomy, as a classification system will be designed on faceted approach base, which could ease the analyzing and retrieval processes.

Regarding building the base of the taxonomy, this project will follow the Literary Warrant technique. The ontology, will involve not just results of subject analysis, but additional information about the related to the sources itself, which called Background Knowledge.

Finally, the instances of this ontology will be limited in a defined sample represents the component of what intended by filmed materials.

## 5 Motivation of this Research

Although there has been a considerable scientific contribution in the field of the ontology literature, there have been no academic efforts toward availing of the experience of library science, particularly classification science. It is widely known that librarianship has a long history in organizing the human knowledge since ancient world. Furthermore, there are many efforts have been done in this field in forming various schemes as Dewey Decimal Classification (DDC) scheme, Colon classification scheme by Ranganathan, or Library of Congress Classification Scheme, which were the main tools for organizing, storing, and retrieve the intellectual production represented mainly in books or paper forms. In the fact, ontologies have developed beyond the fields of Philosophy, Library science, and knowledge representation (Mcguinness, 2003). The lack of communication between multiple disciplines who tackle this topic (Soergel, 1999) affect the collaboration toward developing the area together. For instance, there is obvious terminology confusion between terms such as Classification, Ontology, Taxonomy, Thesauri, and Dictionary...etc.(Soergel, 1999;Choi, 2008b;Rees, 2003) despite of clarity of these terms in library science . Hence, it would be worthwhile trying to close the gap between these two disciplines from the library science point of view to enable each to benefit from each other.

On the other hand, it would be a contribution in enhancing the recognition of cinema as a source of information not merely a tool for pleasure or entertainment. Thus, this will open widely a gate of a new source of experiences, experiments, and a huge amount of knowledge, which could be considered as tacit knowledge, and arise it through represent it in to ease access it and use it as a new source of knowledge.

From ontology aspect, it would be contribution in building an ontology represents the subjects of filmed materials which may consider as a tool can be used to retrieve these materials semantically.

## 6 Overview of the Proposed Approach

The approach we will be taking to develop our ontology is heavily influenced by the Methontology development method (Gomez-Perez et al., 2004;García, 2006;Oscar et al., 2005). The key steps in this approach are summarized below:

### The first stage: Acquisition

1. Sample collection  
In this step the researcher will try to collect data of balanced sample of the filmed materials represent the community to the maximum.
2. Subjects identification  
Identify the subjects contained in these materials

### The second stage: Conceptualization

1. Building a glossary of terms
2. Classify these terms into one or more taxonomies.
3. Defining the binary relations between concepts.
4. Building the dictionary of concepts.
5. Defining binary relations in detail.
6. Defining instances' attributes in details.
7. Defining classes attributes in details.
8. Defining the constancies in details and construct a constant table.
9. Describing the formal axioms.
10. Defining the rules.
11. Introducing the instances details.

### The third stage: Evaluation

This stage involves the following three aspects according to the consistency, completeness and conciseness criteria:

1. Ontology verification, in terms of the ontology being free of errors.
2. Ontology validation, in terms of whether the ontology will be represents the real world.
3. Ontology assessment, by the judgments from the end users point of view.



## 7 Previous Works

There is no doubt that there are many efforts related to this research project topic. Through the research, it has become clear that there are no previous works describing the construction of ontology for the filmed materials based on the subjects implied in these materials, in spite of some efforts that have been made in the field of ontology that tackled the multimedia retrieval.

The main work which can be mentioned here, The movie ontology (MO) “aims to provide a controlled vocabulary to semantically describe movie related concepts such as Movie, Genre, Director, Actor and individuals” ([Http://Www.Movieontology.Org/](http://www.Movieontology.Org/), 2010). MO developed, adapted and maintained by the Department of Informatics at the University of Zurich.

There is another try to build an ontology for movies, whose aim is to design a knowledge representation for cinema descriptions (Tu, 2007).

Although, these two projects have dealt with movie ontology, they both used outward elements to describe the components of these materials. To the knowledge of the researcher, nobody has tried to tackle the substantive content in such material which is the subject.

## 8 References

- Brachman, R. & Levesque, H. (2004) *Knowledge representation and reasoning* Amsterdam Boston, Morgan Kaufmann.
- Broughton, V. (2006) The need for a faceted classification as the basis of all methods of information retrieval. *Aslib Proceedings*, 58, 49-72.
- Cantoni, F., et al. (2002) SOCIO-TECHNOLOGICAL ASPECTS OF KNOWLEDGE DISCLOSURE:THE RELEVANCE OF STORYTELLING IN KNOWLEDGE-ECOLOGY BASED ORGANIZATIONS. *ICIS The Xth European Conference on Information Systems Information Systems and the Future of the Digital Economy*. Gdansk, Poland.
- Carbonaro, A. (2008) Ontology-based Video Retrieval in a Semantic-based Learning Environment. *Journal of e-Learning and Knowledge Society - English Version*, Vol 4, 203-212.
- Chauvel, D. & Despres, C. (c2000) *Knowledge horizons the present and the promise of knowledge management*, Oxford, Butterworth-Heinemann.
- Choi, Y. (2008a) Making Faceted Classification more acceptable on the Web: A comparison of Faceted Classification and ontologies.
- Choi, Y. (2008b) Making Faceted Classification more acceptable on the Web: A comparison of Faceted Classification and ontologies. *Proceedings of the American Society for Information Science and Technology*, 45, 1-6.
- Christopher, B. (2004) Knowledge Representation with Ontologies: The Present and Future. IN Kieron, O. H., Steve, F., Yorick, W., Enrico, F., Mark, A. M., Jeremy, E. & Simon Buckingham, S. (Eds.) *IEEE Intelligent Systems*.
- Dagobert, S. (1999) The rise of ontologies or the reinvention of classification.
- Davenport, T. H. & Grover, V. (2001) General Perspectives on Knowledge Management: Fostering a Research Agenda. *Journal of Management Information Systems*, 18, 5-22.
- Davenport, T. H. & Prusak, L. (1997) *Working knowledge : how organizations manage what they know* Boston, Mass., Harvard Business School.
- Davis, R., et al. (1993) What is a Knowledge Representation? *AI Magazine*, 14, 17-33.
- García, R. (2006) A Semantic Web Approach to Digital Rights Management (PhD thesis). *Technologies Department*. Barcelona, Spain, Universitat Pompeu Fabra.
- Gomez-Perez, A., et al. (2004) *Ontological Engineering : with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. First Edition (Advanced Information and Knowledge Processing)*, Springer.
- [Http://Www.Movieontology.Org/](http://www.Movieontology.Org/) (2010) The movie ontology MO to semantically describe movies.
- Mcguinness, D. L. (2003) Ontologies Come of Age. *The Semantic Web: Why, What, and How*. MIT Press.
- McLellan, H. (2002) Introduction to Corporate Storytelling. <http://tech-head.com/cstory1.htm>.
- Nonaka, I. & Takeuchi, H. (1995) *The knowledge-creating company : how Japanese companies create the dynamics of innovation / Ikujiro Nonaka and Hirotaka Takeuchi*, New York :, Oxford University Press.

- Oscar, C., et al. (2005) Building Legal Ontologies with METHONTOLOGY and WebODE. *Law and the Semantic Web*.
- Passant, A. (2007) Using Ontologies to Strengthen Folksonomies and Enrich Information Retrieval in Weblogs : Theoretical background and corporate usecase. *ICWSM*. Colorado, <http://www.icwsml.org/papers/paper15.html>.
- Plato Theaetetus. <http://onlinebooks.library.upenn.edu/webbin/gutbook/lookup?num=1726>.
- Prieto-Diaz, R. (2003) A faceted approach to building ontologies. *Information Reuse and Integration, 2003. IRI 2003. IEEE International Conference on*.
- Rees, R. V. (2003) Clarity in the usage of the terms ontology, taxonomy and classification. *CIB W78's 20th International Conference on Construction IT, Construction IT Bridging the Distance*. Waiheke Island, New Zealand.
- Ren, F. & Bracewell, D. B. (2009) Advanced Information Retrieval. *Electronic Notes in Theoretical Computer Science*, 225, 303-317.
- Seremeti, L. & Kameas, A. (2007) Multimedia ontologies. *Proceedings of the 3rd international conference on Mobile multimedia communications*. Nafpaktos, Greece, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- Soergel, D. (1999) The rise of ontologies or the reinvention of classification. *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE*., 50, 1119-1120.
- Sole, D. & Wilson, D. G. (2002) Storytelling in Organizations: The power and traps of using stories to share knowledge in organizations. *LILA Harvard University*.
- Tjondronegoro, D. & Spink, A. (2008) Web search engine multimedia functionality. *Information Processing & Management*, 44, 340-357.
- Tu, A. (2007) Building an Ontology for Movies [www.andrew.cmu.edu/user/atu/cs/building\\_an\\_ontology\\_for\\_movies.doc](http://www.andrew.cmu.edu/user/atu/cs/building_an_ontology_for_movies.doc).
- Vickery, B. C. (1968) *Faceted classification: a guide to construction and use of special schemes; prepared by B. C. Vickery for the Classification Research Group*, London, Aslib.
- Welty, C. A. & Jenkins, J. (1999) Formal ontology for subject. *Data & Knowledge Engineering*, 31, 155-181.
- Zhu, H. & Madnick, S. E. (2006) A Lightweight Ontology Approach to Scalable Interoperability. SSRN.