OntoMedia: An Ontology for the Representation of Heterogeneous Media

Michael O. Jewell
Image, Speech and Intelligent Systems Research Group
School of Electronics and Computer Science University of Southampton United Kingdom SO17 1BJ
moj@ecs.soton.ac.uk

Faith Lawrence
Intelligence, Agents, Multimedia Group
School of Electronics and Computer Science University of Southampton United Kingdom SO17 1BJ
kf03r@ecs.soton.ac.uk

Mischa M. Tuf eld
Intelligence, Agents, Multimedia Group
School of Electronics and Computer Science University of Southampton United Kingdom SO17 1BJ
mmt04r@ecs.soton.ac.uk

ABSTRACT

With the emergence of the Semantic Web, a shared vocabulary is necessary to annotate the vast collection of heterogeneous media already in existence. The OntoMedia ontology aims to provide a meaningful set of relationships which may enable this process, while being suited to mapping and extension. In this paper we outline the salient features of our ontology as well as initial applications and comparisons to existing technologies.

Categories and Subject Descriptors

H.3.1 [**Content Analysis and Indexing**]: Abstracting Methods

General Terms

Standardization

Keywords

Ontology, Multimedia, Semantic Web, OntoMedia

1. INTRODUCTION

A key problem faced by the information management society is that of information overload. The ever growing, evolving nature of the World Wide Web is not making information retrieval any easier. The web is populated with an unmanageable amount of heterogeneous data, in forms as diverse as photos, sound, video, and text. The shift in the consumer electronics market from analogue to the digital recording medium and the mass-market, low cost nature of the industry has brought about this rich corpus of digital media.

In this paper we propose a vocabulary, in the form of an ontology[10] for the annotation of multimedia documents. The resulting annotation, or markup, applied to a piece of media will provide, a semantically rich description, in the form of metadata, in a machine-readable format. One of the motivations behind making such knowledge explicit, as apposed to simple keyword labeling common to the current Web, is to capture and describe knowledge that is implicit in the context of the given media unit[13].

The Semantic Web[4] defines the necessary relational models for describing web resources with context independent standards, such as the Resource Description Framework[16] (RDF). RDF is currently used to describe vast amounts of different data, and is defined to be applicable to the description of multimedia artifacts. The vocabulary presented to annotate the multimedia items is specified as an ontology, where all of the definitions are presented with respect to each other and their semantic relationships are defined explicitly[17]. The ontology should be used as a common vocabulary of definitions relating to multimedia markup, allowing for knowledge to be shared and communicated between people and machines, to facilitate reasoning over this shared conceptualisation upon heterogeneous multimedia.

The currently emerging field of Semantic Web technologies is challenging the manner that authors publish information, from the classic method of developing a document that is solely intended to convey a message to a human reader, to the publishing of elements of raw knowledge in the form of annotated media items that are linked together in a structured and meaningful manner. This new paradigm of publishing would allow for content-based multimedia to be exploited in aid of generating on the fly narratives from the available knowledge elements in manner best suited to the user's profile[5].

Advances in the techniques and tools for the semantic enrichment of information have allowed for relationships in data to be made explicit in knowledge rich domains. One of the cornerstones of developing Semantic Web applications is the development of ontologies. Advances in ontological driven multimedia presentation assembly from semantically enabled knowledge elements have been reported in recent work[9][15][5]. This assembly of structured multimedia presentation from a collection of knowledge elements will only be enriched by the wide adoption of the Semantic Web, and the subsequent availability of more annotated knowledge nuggets. Summarisation[2], document generation systems[20][8]], and system that aid biography tailoring[12] have benefited from the proposed method of semantic enrichment.

The metadata produced by annotating with respect to a shared ontological vocabulary will allow for searching and navigation by concept, an example of such a system is Sculpteur[1]. Sculpteur is a system that allows users to search and navigate semantically enriched museum multimedia; the system also demonstrates how metadata, from a shared vocabulary CIDOC (see below), can be used to search across media stored in heterogeneous data stores.

The annotation of media into an ontological form has been investigated from a variety of angles by prior researchers, most significantly by Lagoze and Hunter's ABC Ontology[14]. This was designed primarily for the cataloguing community, with a focus on factual information representation, such as object provenance and rights management. ABC and OntoMedia share the separation of entity and temporal classes, whilst the OntoMedia ontology augments this core with classes that allow for specialization to fiction. Furthermore, Hunter proposes a technique to represent MPEG-7 using a DAML+OIL representation, whereas we reference segments of the source media using VLit location specifiers (see Section 3.1).

The cultural significance of such an ontology is further emphasised by the CIDOC Conceptual Reference Model (CRM), which was created as a "semantic approach to integrated access"[7] for cultural heritage data. By providing a conceptual basis that can be used for automated mapping the CIDOC CRM acts as a bridging technology between existing data structures and a guide to creating new structures.

The similarity between cultural heritage and fictional content is one of underlying components. Both are concerned with people and events, the only difference being the type of evidence that exists of those manifestations and where they were believed to have taken place. The fictional aspect of the historical narrative has been commented on by figures as diverse as Plato and Churchill. The historical aspect of fictional narrative is more often at the discretion of followers of a particular work. However, ignoring the fact that it is an imaginary history, works of fiction still tell a story. It is therefore not surprising that the basic top-level structures that we discovered worked best for describing the contents of a fictional account were very similar to those of the CIDOC CRM.

2. MOTIVATION

The creation of the OntoMedia ontology entailed three requirements; namely the requirement to be able to annotate cultural data, textual fiction, and film. The CIDOC CRM already focuses on the first of these, and we are assisting with the development of the ontology. However, the latter two are less formalized, and as such acted as the direction for the development of OntoMedia.

Amateur fiction is created, shared and archived in vast amounts, all within the digital environment of the Internet. As such this was a significant motivating factor to the creation of the OntoMedia ontology. The subproject of the OntoMedia scope is concerned with the social networks and webs of trust that exists between the authors and consumers of what, if taken as a whole, is arguably the largest and one of the oldest distributed electronic libraries currently in existence.

Due to both the amount and nature of the material and its hobbyist origins the current systems of storage, access and retrieval are immensely variable and frequently contentious. A group created to allow interaction with some of those involved in this community and create a taxonomy of terminology and meaning quickly discovered that not only were many of the definitions vague but many were fluid and gained specific significance or meaning within one community that was lost or transformed in another. Only interaction within any of the communities allowed the user to pick up the most common meaning for the term within that group. This could be very confusing for those people coming into the group either from another community or from outside.

One of the most striking examples of this was the use of the word *gen* to describe a work of fiction. Depending on the context it could either be interpreted to indicate a story with no romantic interests or one in which the romantic interests were of a level relative to the source material which inspired the piece. In addition to metadata to describe the overall content of the piece, it is often necessary to include metadata which indicates the presence of spoilers within specific events of the story.

As a result of this shifting vocabulary it became clear that marking up these works to a level of detail that would engender a semantic application would require the metadata to exist at a level beneath that of the community specific concepts. While this had the potential of requiring more detail than would otherwise be needed it also offered the possibility of reader-defined and personalized lexicons of terms, thus increasing the usability of the system and lowering the level of expertise that would be required of a new user. It also made it possible to map the contents to current Internet content selection systems such as the World Wide Web PICS specification[18].

While literature provides a textual version of a situation, film provides visual, audible, and sometimes textual (in the cases of scripts or closed captioning). As such, this was one of the most complex candidates considered for the use of OntoMedia. As well as following the timeline/entity design inherent in multimedia, it also contains a wide variety of extra information. The ability to annotate this diverse collection of media formats using a single ontology would allow for ease of automatic markup. For example, the marked version of an audio representation could be validated against the marked version of the textual. Furthermore, the different media provide alternate viewpoints, with the video and audio versions containing the director's interpretation of the script.

While we are focusing on these three modes, an overall motivation to the OntoMedia project is to provide an ontology which can represent any of the heterogeneous media which

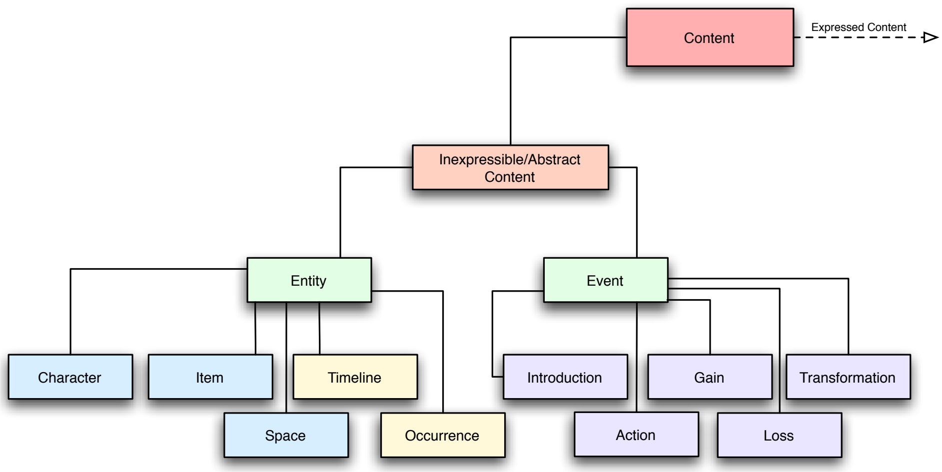

Figure 1: The structure of the OntoMedia ontology

are prevalent on the internet. As such, the semantic information contained within these media will be available on a cross-platform, multi-modal, basis.

3. THE ONTOLOGY

At the center of the OntoMedia design are entities and events. These two classes represent the elements present in a media and, where appropriate, the situations in which they are present. This high level of abstraction is capable of encompassing a wide variety of media, whether factual or fictional, and is not restricted to modelling interactions between people. Figure 1 shows an overview of the OntoMedia classes, with the division between entities and events apparent.

3.1 Event Modelling

An event within a given piece of media consists of one or many interactions between participating entities, and may be instantaneous (happening at a specific moment) or continuous (happening throughout a set period). As shown in Figure 2 there are three core events, namely `Gain` and `Loss` (e.g. a character learning some information or losing money), and `Transformation` (e.g a character becoming older). The classes of events are based on those contained in the 'typical' story, as discussed by Bal[3] and Chatman[6]. While these may seem simplistic initially, they may be extended - so `Loss` can be extended to `Destruction` or `Betrayal` (the loss of a bond) and `Transformation` can contain `Travel` (locational transformation). The formal definition of an event in OntoMedia is "an interaction between one or more entities during which zero or more traits of those entities are modified and/or a new entity is created".

Furthermore, each event may have preconditions and postconditions which are required for the event to take place or be judged as complete. For example, the event in which a character loses money requires that the character has money initially. This information, though not so useful for inference, is ideal for the markup process as it is then possible to ensure that events only occur when it is possible for them to happen.

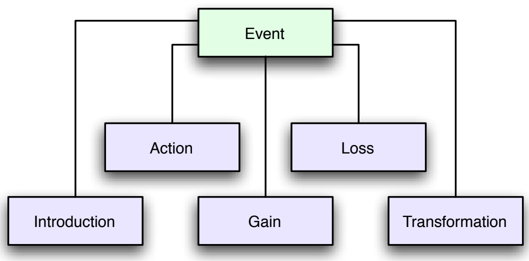

Figure 2: The OntoMedia event hierarchy

To allow for the modeling of event chains, such as a situation that is likely to arise as a result of another, OntoMedia provides ordering properties. An event may cause, or be caused by, another event (not necessarily preceding). Note, however, that these orderings do not impose any timing information on the event objects, as this extra layer is provided by occurrence representations. Occurrences place events into a temporal context. This allows for the same event to occur multiple times in the same media, possibly in several timelines. An occurrence is a straightforward class, with a Terminus Ante Quem (TAQ) providing the location specifier of the beginning within the medium, and a Terminus Post Quem (TPQ) providing the location specifier of the ending. Location specifiers provide a means to reference portions of media in an extensible and multimodal manner, and hence allow for events to occur within media which are either spatial (such as photographs or comics) or temporal (such as audio or video). The occurrence also contains references to the Event which is occurring, and the timeline in which it occurs.

3.2 Entity Modelling

The entities in a media, which make up the other half of the core OntoMedia classes, represent the items or concepts which participate within the contained events. As such they include physical entities, such as characters or props, and

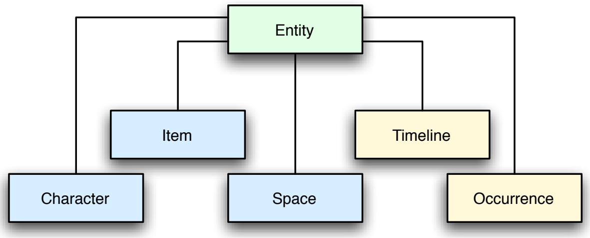

Figure 3: The OntoMedia entity hierarchy

abstract entities, such as language or culture (see Figure 3).

The base-level entity provides a few key properties which are inherited by both the abstract and physical subclasses. These include container information, allowing for one entity to be contained by another, location information, which refers to a custom location ontology, and a collection of 'traits'. These are fundamental to the OntoMedia representation, as they embody the characteristics and properties of entities within the media.

OntoMedia defines traits which cover the most common attributes which were found to occur in fictional media. These include personal information, such as age and faith, physical information, such as build and distinguishing marks, and state-based attributes, such as being and form. As mentioned previously, traits may only be altered or added as the by-product of an event, so it is feasible to denote a character's physical appearance altering as the result of a transformation event.

The final, and possibly most powerful, trait within the OntoMedia description is that of 'motivation'. This defines the state that the entity is aiming to achieve in order to gain fulfilment. The trait contains zero or more event instances as well as zero or more entity instances which represent the goal state which the subject aims to realize. By analysing the state of the medium with regards the motivation of a character, it is therefore possible to determine whether the entity has fulfilled its goals by the culmination of the piece.

3.3 Extensibility

With the aim of modeling the contents of fiction, regardless of media format, it was necessary to allow for extensibility of the framework. For example, OntoMedia incorporates the location ontology created for the Signage project[17] to provide a basic spatial model. This ontology provides the requisite level of detail which arose from example cases that were analysed during the course of the design process.

Focusing on the needs raised by those examples we extended this ontology where it was insufficient for our purposes. The top division between `Enclosed_Space` and `Unenclosed_Space` proved suitable with only the addition of a `Surface_Space` class necessary to cover the those cases where the intended area was two dimensional rather than three. Further classes were also added under `Unenclosed_Space` to match the granularity that existed within the `Enclosed_Space` tree. The extended version of this section of the ontology can be seen in Figure 4.

The examples used to test the integration of the location ontology were taken from both multimedia sources and literature. However, because the requirements and restrictions were deliberately set at the `ontomedia:Space` level a more or less detailed spatial model could be substituted if required. This is illustrated more clearly in the case of the humanoid body parts ontology. The ontology itself subclasses both `ontomedia:Physical_Item` and `ontomedia:Surface_Space` and was envisioned to be used primarily in conjunction with the `ontomedia:Character_Description` and `ontomedia:Distinguishing_Mark` traits. A work of fiction with a medical setting might require a much more accurate model of the human body while other works might require more detail in certain areas of anatomy.

The location and body parts ontologies were always constructed separately to the OntoMedia model, whereas the profession model was originally created as part of the main ontology. Having tested this arrangement it quickly became apparent that it was not ideal. While having a generic profession ontology was useful as a time saving measure the overlap between the professions required by any two works of fiction was so small and the range of possible professions so large that even the basic break-down into profession types was too cumbersome for the main ontology.

This concept of creating reusable models that could be included when necessary was one that was explored furthere with our use of the `ontomedia:Context` class.

3.4 Contexts

The `ontomedia:Context` class was created to separate the many different versions of the same entity that may exist. This is a particular issue when considering the contents of fiction, especially when those works have been reinterpreted across media, within the same work or after a period of time. As these different interpretations may be physically distinctive, for example when a character is portrayed by different actors or given different personality traits it becomes necessary to recognise that there are occasions when their differences are as important as their similarities.

Examples of this can be seen in almost every movie adaptation of a book. For example, the recent Lord of the Rings movie or the transformation of 'We Can Remember It For You Wholesale' to the movie 'Total Recall'. In the first case the character of Faramir (as portrayed in the movie by David Wenham) was both physically and emotionally different to the character described in the book. In the second, the hero of the short story 'Douglas Quail' becomes 'Douglas Quaid'.

Interaction with the community showed that they frequently distinguished between these different representations as well as creating their own. As this separation was significant to them it was also important to be able to model that distinction and allow the metadata to reflect it in the annotation for the purposes of reducing ambiguity in search and retrieval.

The use of distinct contexts allows us to differentiate both between different representations of the same fictional character and between real people, fictionalisations of real people

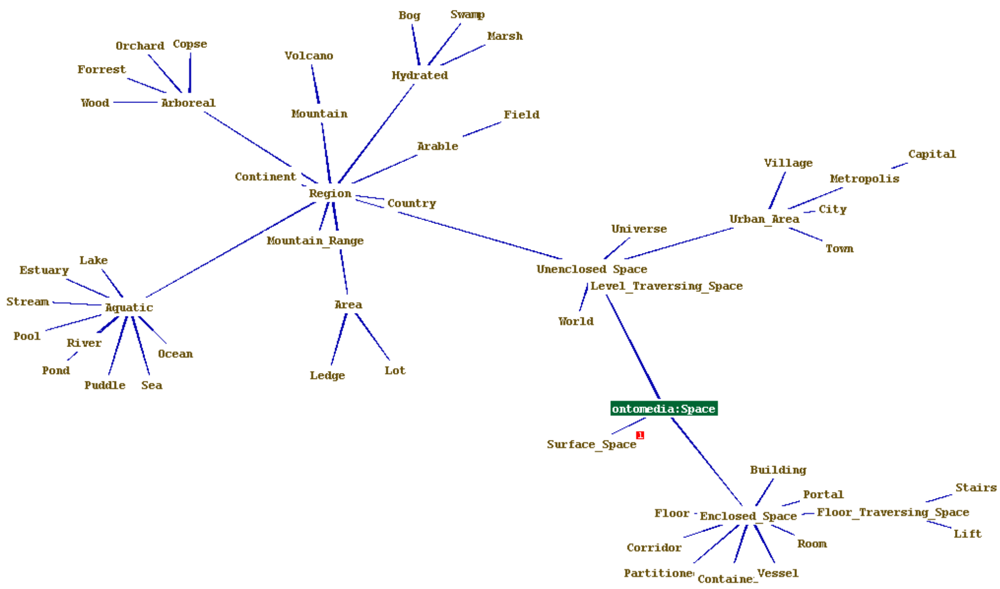

Figure 4: The OntoMedia extended location ontology

and other fictional characters. A factual context was defined as a top level instance to allow entities to be defined as existing in the real world.

The relationship between these different types of intercontextual constructs cannot be defined as simply stating they are the same entity. In some situations a simple 'is_a' relationship is appropriate but in others a more complex link exists and declaring that the two are the same would be potentially misleading and inaccurate. To better represent these situations we allowed looser links to be defined to allow a greater range of inferences to be made about entities.

4. PRELIMINARY RESULTS

Initial tests of the OntoMedia framework focused on the aforementioned short story 'We Can Remember It For You Wholesale'. As well as containing interesting characterizations and plot direction, the story was also chosen for its changing timelines (some sections are written as the dreams of the main character) and cross-media possibilities (due to its movie adaptation).

Using an annotated version of the story imported into a triple store[11], several SeRQL queries were carried out. These included the retrieval of events featuring specific characters, the automatic construction of a cast list, the identification of all characters of a specified gender, and the location of 'key' scenes based on spoiler-specific metadata. All of these tests were successfully performed using the OntoMedia representation.

Character	Weighted Score
Douglas Quail	105
Interplan	15
Kirsten Quail	10
Martian Politician	7
Alien Invaders	6
McClain	2
Keeler	2
Lowe	2
Shrink	2
SS1	1

Table 1: The characters in the story 'We Can Remember It For You Wholesale', with scores representing their involvement in the storyline.

Building on these already encouraging initial results, a more complex problem was designed. In order to identify important events within a media, it was felt that it would first be necessary to discover presence of key characters. To achieve this, three queries were constructed; the first to enumerate the occasions in which a character was the *subject* of an event, the second to enumerate those cases where a character was the *object*, and the third to identify characters present in a scene but not actively participating. Once attained, these totals where weighted to reflect the significance of the characters' presence in the events, with subject cases multiplied by 4, object cases by 2, and 'involved' cases remaining unchanged.

The results of this calculation applied to 'We Can Remember It For You Wholesale' can be seen in Table 1. It is immediately evident from these figures that Douglas Quail is the key character in the story, which is logical as the plot unfolds from his point of view, but it is also straightforward to identify the other major characters (Interplan, Kirsten Quail, the Martian Politician, and the Alien Invaders) as well as those deemed secondary (McClain, Keeler, et al).

While this example provides a powerful use in itself, it will also be possible to examine the differences between the short story and the movie adaptation. Both media may be annotated using the same OntoMedia structures, so a comparison of character prominence can be carried out by applying the same queries to both media types.

5. FUTURE WORK

The possibility of mapping the OntoMedia ontology to the CIDOC CRM system opens up a number of possibilities in the heritage field. Work is already being done to harmonize the CIDOC CRM and the Functional Requirements for Bibliographic Record[19] (FRBR) conceptual model which was created by the International Federation of Library Associations and Institutions. While the FRBR model has been used to prototype a 'fiction finder' application it works solely at the bibliographic level of information. One can search within a specific genre or for a known author but not on a type of action that is contained or a particular character. The addition of this type of metadata to cultural heritage texts, especially when linked by the CIDOC CRM to non-literary artifacts or information, could expose previously unnoticed links and provide a new tool for researchers.

Beyond the factual context, the OntoMedia model does not distinguish between real and fictional. As such it can be used to annotate and describe the mythological histories that exist in both written and oral tradition. Comparative mythology would benefit significantly from the ability to search for myths of a particular type or ones in which certain events occurred. This information can then be combined with external sources such as historical events and trade patterns to provide an integrated view of the conditions that existed in the period under investigation, as well as the influences that would have existed from and between the social, economic, and religious elements of that time.

From the film perspective, work is underway to provide techniques for automatic annotation into the OntoMedia vocabulary. This will utilise feature extraction techniques to augment the existing video and audio data with the knowledge contained therein. The portability of the OntoMedia framework will allow for crossmedia analysis and information retrieval in areas such as media searching, automatic rating, and genre classification. Work is already underway to build a corpus of annotated media, and this will gain momentum as automatic techniques are developed.

6. CONCLUSIONS

The OntoMedia ontology provides a powerful annotation vocabulary which is both multimodal and extensible. As well as building upon the work done in the ABC and CIDOC projects, the capability for handling more elaborate media formats and content. The use of location specifiers provides possible ties into photography, audio, and film, as well as textual information.

Core to the OntoMedia philosophy is the ability to extend a generic framework in such a way to accommodate requirements for media markup. In this paper we have highlighted the use of metadata specific to fiction, but it is equally possible to represent the metadata for factual media. Furthermore, the constructs provided for motivation and trait description may be employed both for annotation and analysis.

With the emergence of the Semantic Web, it is increasingly important for the annotation of data into machine-readable formats. The OntoMedia ontology provides a shared conceptualization of this domain which will act as an enabling technology for this evolution of the web.

7. ACKNOWLEDGMENTS

This work is partially supported under the Advanced Knowledge Technologies (AKT) Interdisciplinary Research Collaboration (IRC), which is sponsored by the UK Engineering and Physical Sciences Research Council under grant number GR/N15764/01.

8. ADDITIONAL AUTHORS

Additional authors: Adam Prügel-Bennett, David E. Millard, Mark S. Nixon, monica schraefel, Nigel R. Shadbolt

9. REFERENCES

[1] M. Addis, P. Lewis, K. Martinez, J. Stevenson, and F. Giorgini. New ways to search, navigate and use multimedia museum collections over the web. In *In Proceedings of Museums and the Web 2005*, 2005.

[2] H. Alani, S. Kim, D. E. Millard, M. J. Weal, P. H. Lewis, W. Hall, and N. R. Shadbolt. Automatic ontology-based knowledge extraction from web documents. *IEE Intelligent Systems*, 18(1):14–21, January 2003.

[3] M. Bal. *Introduction to the theory of Narrative.* University of Toronto Press, Toronto:Canada, 2nd edition, 1997.

[4] T. Berners-Lee. *Weaving the Web.* Orion Business Books, 1999.

[5] I. M. Bilasco, J. Gensel, and M. Villanova-Oliver. Stamp: A model for generating adaptable multimedia presentations. *Int. Journal Multimedia and Applications*, 25(3):361–375, 2005.

[6] S. Chatman. *Story and Discourse.* Cornell University Press, New York, 1978.

[7] M. Doerr, J. Hunter, and C. Lagoze. Towards a core ontology for information integration. *Journal of Digital Information*, 4(1), April 2003.

[8] J. Domingue, E. Motta, S. B. Shum, and S. Vargas-Vera. Ontology-driven document enrichment within communities of practice. In *In Proceedings of 1st Int. Conference on Knowledge Capture*, pages 30–37. International Conference on Knowledge Capture, 2001.

[9] J. Geurts, S. Bocconi, J. van Ossenbruggen, and L. Hardman. Towards ontological-driven discourse: From semantic graphs to multimedia presentations. CWI Technical Report INS-R0305, 2003.

[10] T. Gruber. Cidoc conceptual reference model committee. In *CIDOC Conceptual Reference Model Committee Meeting Proceedings*, March 2003.

[11] S. Harris and N. Gibbins. 3store: Efficient bulk rdf storage. In *In Proceedings of 1st International Workshop on Practical and Scalable Semantic Systems*, pages 1–15. (PSSS'03), 2003.

[12] S. Kim, H. Alani, D. E. Millard, M. J. Weal, P. H. Lewis, W. Hall, and N. R. Shadbolt. Generating tailored biographies with automatically annotated fragments from the web. In *In Proceedings of Workshop on Semantic Authoring , Annotation, and Knowledge Markup*, pages 1–6. 15th European Conference on AI (ECAI), 2002.

[13] R. Klamma, M. Spaniol, and M. Jarke. Mecca: Hypermedia capturing of collaborative scientific discourses about movies informing science. *Int. Journal of an Emerging Discipline: Special Series on Issues in Informing Clients using Multimedia Communications*, 8:3–38, 2005.

[14] C. Lagoze and J. Hunter. The ABC ontology and model. *Journal of Digital Information*, 2(2), November 2001.

[15] S. Little, J. Geurts, and J. Hunter. Dynamic generation of intelligent multimedia presentations through semantic inferencing. In *In Proceedings of 6th European Conf on Research and Advanced Techniques for Digital Libraries*, pages 159–189. Springer - ECDL, 2002.

[16] F. Manola and E. Miller. Rdf primer: W3c recommendation. http://www.w3.org/TR/rdf-primer/, 2004.

[17] I. Millard, D. D. Roure, and N. R. Shabolt. The use of ontologies in contextually aware environments. In *In Proceedings of First International Workshop on Advanced Context*, pages 42–47, 2004.

[18] J. Miller, P. Resnick, and D. Singer. Rating services and ratings systems (and their machine readable descriptions. http://www.w3.org/TR/REC-PICS-services, 1996.

[19] K. G. Saur. Functional requirements of bibliographic records: final report. Technical Report, IFLA Study Group on the Functional Requirements of Bibliographic Records, Munich, 1998.

[20] N. D. Silva and P. Henderson. Narrative support for technical documents: Formalising rhetorical structure theory. In *In Proceedings of Enterprise Information Systems (In Press)*. ICEIS, 2005.