A comparative analysis of ontological techniques in supporting online

communities

JOHN MCAULEY, EOIN KILFEATHER

Digital Media Centre, Dublin Institute of Technology, Ireland, john@dmc.dit.ie

ABSTRACT

In the early years of the Internet, online communities (OC) were considered primarily a social phenomenon. However, the last five years have seen a paradigm shift in the methodologies used to support virtual communities. The commercial sector had been quick to recognise the benefits of adding more depth and vitality to their sites through the inclusion and encouragement of user created content. The effects of a strong commitment and loyal membership towards sustaining the community can significantly add to their success. However, with the advent of knowledge management as a recognised and increasingly practised discipline, new techniques and methodologies have been developed to enhance online collaboration and content management. In this context narrative is still understood to be the principle mode of knowledge transfer within an online environment. In design and use it seems that online cultural heritage discourse is necessarily mediated by this form. Within this context, this paper describes four separate approaches undertaken as part of the CIPHER project in providing intelligent support to online communities within the Cultural Heritage sector.

1. INTRODUCTION

This paper presents a comparative analysis of existing techniques being employed in the development of content management and community supporting platforms and their applicability in the heritage domain. The basis for the paper comes from the research accomplished in the CIPEHR (Communities of Interest Promoting Heritage of European Regions) project. This was a two and half year EU project carried out under the theme 'heritage for all' in the IST 5th framework research programme. The CIPHER approach involved the development of innovative technologies to foster online communities and aid in the exploration of online Cultural Heritage (CH) content. Rather than adopt a monolithic or single stranded approach the CIPHER project experimented with different support mechanisms for online communities but all with a common organisational metaphor of narrative. This paper gives a concise introduction to the concepts of ontology and how ontologies can be used to support CH narratives. A brief summary of the CIPHER project and a description of the four forums developed and trialed during the project's lifetime is also given. The individual approaches to forum design are highlighted with particular attention paid to the application of knowledge support and awareness of narrative concepts. The paper concludes with an introduction to the TARCHNA (Towards Archaeological Heritage New Accessibility) project, in which the authors intend to utilise many of the concepts developed during the research carried out on CIPHER.

2. THE PROVISION OF INTELLIGENT COMMUNITY SUPPORT

Currently, artificial intelligence, knowledge management and the semantic web initiatives are firmly rooted in the belief that the conceptual representations and descriptive power of ontologies greatly aid in the support of more intelligent computer applications. For the purposes of this paper ontology will be defined as 'a formal, explicit specification of a shared conceptualization' [1]. However, ambiguity still exists regarding what types of conceptualisation deserve the label 'ontology'. Within this context, McGuinness defines an ontology spectrum (figure 1) in describing a typology of ontologies in order of their descriptive power [2]. It can be seen from the figure 1 that ontology classification can be represented by a simple model such as a glossary of terms (Non-formal), through to a more complete representations such as frame-based models (formal) shown on the right hand side of the spectrum. The red line which dissects the spectrum represents the crossing from informal and semiformal ontologies to more formal types of computational representations. McGuinness' spectrum is used to contrast both the implementation of ontology models within a community environment and the descriptive power provided by the separate approaches as described in section 3.

3. THE CIPHER APPROACH

The CIPHER project aimed to support communities of interest in their exploration of regional heritage through online cultural heritage (CH) forums. A CH forum may be thought of as an 'online space where people can participate and learn through accessing and contributing to a range of heritage resources organised around a common theme' [3]. From the outset the project set out some approaches which differentiated the CIPHER cultural heritage forums from other cultural

portal approaches. Members of the forum are viewed as active participants rather than a passive audience. In this way a CH forum is different than a portal website, as the user actively explores CH content rather than passively consuming static narratives published by heritage professionals [4]. Six European partners were involved in the research, they were the Dublin Institute of Technology and the Discovery Programme from Ireland, the Knowledge Media Institute from the UK, the University of Art and Design Helsinki based in Finland, the Czech Technical University and Regionales Information System (RiS) GmbH from Austria.

The six partners were involved in the development of the four CH forums organised around some specific local themes, these were: Irish cultural and natural heritage, Nordic heritage through storytelling and historical artefacts, the shared heritage of central Europe and the tradition of technology innovation in south central England. Although, each partner collaborated within the project, individual approaches were taken in the design and implementation of each forum. Narrative was given a primary position as the key means of knowledge transfer within an online environment. Within this context the forums aimed to exploit both the pervasive and pedagogical nature of narrative through the introduction of new and innovative dynamic methods of both narrative creation and exploration. However, in its realisation, each forum took separate approaches in supporting the community exploration of CH content.

3.1 THE IRISH FORUM

The Irish Forum (explorer.ie) as described in [5], [6], [7] was developed by the Dublin Institute of Technology in association with Discovery Programme and gives online access to a database of Irish archaeology held by the latter. The Irish forum aimed at providing the community with tools that would facilitate a meaningful and coherent story or narrative structure to be developed from existing or newly created content.

3.1.1 DOMAIN CONTENT

It was felt that the impetus to develop websites is often purely technological, rather than addressing the genuine need of the heritage community. Within this context, the explorer forum aimed at containing a body of catalogued archaeological data which could be built upon and a narrative layer formed to help develop a descriptive explanation of that data. The catalogued archaeological data consisted of the NMR (National Monuments Record), a complete list of categorised monument data of archaeological or historical interest for Ireland and Northern Ireland showing geographical positions.

3.1.2 APPLICATION OF KNOWLEDGE SUPPORT

One of the key difficulties facing the designers when opening up remote access to the Irish NMR databases and linking resource discovery was the lack of a properly structured thesaurus of terms for the Irish monuments and archaeological objects. In discussions with the domain experts it was decided to use the forum in a pilot project to explore a manageable section of the Irish CH database and to map these data to a domain ontology. The domain ontology drives the narrative structure of the forum and provides a standardised way of accessing the National Monuments Record. As there is no formal domain ontology representing Irish Archaeology, the problems arose of what type of ontology to use and whether to design a new one from first principles or to adapt an existing ontology.

Both the focus and form of the community contributed to the choice of ontology. It was felt that the needs of a community of interest differed from other online communities. Members joined the group out of a genuine curiosity in their regional heritage and the authoring tools available within the forum should invite content contribution. The use of a highly descriptive property driven ontology could intimidate users when annotating new narratives. This pragmatic question brings the type of community into focus. Knowledge systems are chiefly employed within a corporate environment where annotation is compulsory, when dealing with a community of interest (Col) this is not the case. Cols are more organic and will often grow through a grassroots approach, many users might not see the benefits of the additional annotation overhead and it was felt that this could stifle contribution to the knowledge base.

However, it was also felt that the ontology had to describe the domain accurately providing adequate, appropriate and understandable terminology or 'story concepts' on which the users could develop their narratives. Consulting McGuinness's ontology spectrum (figure 1), a simpler ontological structure was decided upon, where narratives could be related to domain concepts without the arduous annotation of concept properties. As the forum's content is meant for human consumption, it was felt that a thesaurus with it's a hierarchy of terms and additional relationships provided the appropriate domain ontology. From the standpoint of the explorer forum there had been few vocabulary standardisation results in Ireland itself and no immediately suitable ontology of terms existed. However, there were a number of CH standards which had been devised and adopted by bodies in the UK. English Heritage (EH) has developed a series of thesauri in support of major computerization projects. Having previously collaborated with the Discovery Programme in the Forum for Information Standards in Heritage (FISH), English Heritage agreed to the use and extension of archaeological objects and monument types thesauri within the Explorer forum. To specify temporal relations of particular objects and monuments a third vocabulary was add specifying the time periods throughout Irish history.

4. SUPPORT FOR DYNAMIC NARRATIVE AUTHORING

An aim of the forum was to build a narrative layer to accompany the domain content in illustrating Irish heritage. The narrative layer would be produced through the application of the CIPHER knowledge cycle (figure 2). Two narrative components were recognized in providing context and supporting this cycle. The first is a basic narrative unit called a 'story', the second is a broader component called a 'story trail'. From a user's standpoint a story is a piece of text which relates to some concept of cultural heritage while a trail represents a collection of stories juxtaposed in a linear format. Three authoring tools were developed to support the creation of narrative within the forum; they are the *Story Creator* and the *Story Trail Creator*. The tools provide the user with a uniform environment for preparing narratives and their associated annotations. To accompany written narrative a third tool was developed supporting users in the creation of 3D narrative presentations.

The tools support the following functionality:

- Knowledge Acquisition (finding): the tools allow for the insertion and organisation of knowledge in terms of heritage artefacts, story characteristics, etc.
- Interpretation of heritage resources (understanding & organising): The forum supports the interpreting phase of the knowledge cycle through the active annotation of new resources (image, film, sound) to the forum.
- Construction of new stories (composing): The Story Creator allows the user to construct new stories by retrieving relevant concepts and annotated resources from the knowledge base using semantic search criteria.
- Publishing of structured narrative content: When the author is satisfied with the content, the story is published for public consumption on the forum.
- Juxtaposition of narrative: Following the initial trials, the domain experts who had used the system indicated that they felt a dynamic narrative system which attempted to match stories solely through an ontological association was too restrictive. To this end, the *Story Trail Creator* was developed to allow forum members to publish as a linear narrative format by collecting stories under a common theme.

The implementation of both the thesauri and the glossary of time periods delivered additional exploration functionality to the forum. The user could explore stories by their principle concepts through a question driven interface, e.g. Tell me about weapons found around hill-forts during the Iron Age. Furthermore, the thesauri provided a means for 'triangulating' stories linked by similar story concepts. Story concepts allowed stories to be cross connected by different themes and thereby providing different views of the concept by different authors. It was felt that the juxtaposition of stories related to a singular concept provided the reader with a better understanding of the central concept, while mapping concepts through the narrative provided a better understanding of the context for the actual story.

Name	Description
Story understanding (SU)	Understanding stories in terms of the concepts they contain and their relationship to them.
Concept understanding (CU)	Understanding concepts in terms of the how they are used in stories and in what context.
Concept comparison (CCm)	As in CU but comparing across multiple concepts.
Concept connecting (CCn)	Connecting concepts, finding out how they can be related through the stories.
Story mapping (SM)	Mapping story relationships, what occurred when?

Fig. 3 - The five exploration facilities provided by the Explorer forum [9].

4.1 THE NORDIC FORUM

The Nordic forum was developed by the University of Art and Design Helsinki based in Finland in association with Harkko Museum in Raisio, also in Finland. The forum as described in, [10], [11] and [12] uses an innovative technique to produce similarity cluster (SC) representations of its domain content. This offers open and dynamic interpretation of domain narratives to the forum members.

4.1.1 DOMAIN CONTENT

The forum is principally centered on two renowned cultural heritage artefacts, 'A description of the Northern Peoples',

1555 (a book) and the Carta Marina¹ of 1539 (a map). Both were created by Olaus Magnus, the last Catholic bishop of Uppsala, Sweden before his death in 1557. The Carta Marina provided the first comprehensive description of the landscape and people of the Nordic region. In addition the map displays a host of monstrous mythical figures inhabiting the Nordic region. It is generally considered that his 'A description of the Northern peoples' is a broader commentary on the cartographic detail on the map.

4.1.2 APPLICATION OF KNOWLEDGE SUPPORT

As stated in the introduction, each of the CIPHER forums took an innovative approach in developing new ways of delivering Cultural Heritage content to the community. While the explorer forum supplied access to domain content through coherent dynamic narratives driven by the domain ontology, the Nordic forum addressed the provision of knowledge support differently. From this standpoint, it is worth noting the difference in domain content between the two forums. The explorer forum aimed at building a narrative layer on top of a catalogue of archaeological monuments using newly authored narratives. In contrast the Nordic forum aimed at offering new visualization and representational views of the underlying CH content. The CH content consisted of a substantial body of text, already created, which is ideally suited to the application of statistical analysis. The use of spatial vector-based computational techniques in accessing the underlying CH content is independent of pre-determined ontologies and conventional curatorial practices. Within this context, the different approaches rely heavily on the domain content. Statistical analysis works best with a large body of work on which it can draw conclusions, furthermore by using such techniques, the need for, an often lengthy, manual annotation phase of data input is removed.

The use of statistical analysis and semantic recognition offers a unique perspective for visitors to the Nordic forum. Often cultural exhibits are presented from a curator's point of view [13]. However, by applying intelligent statistical approaches to the explanation of artefacts the curatorial onus is taken from the exhibitor. The forum members themselves are left to interpret the exhibit as they wish, by either engaging the underlying content directly, or using the visualisations produced through the forum tools. A variety of statistical and semantic clustering techniques were utilised in the development of the Nordic forum. The Self Organising Map (SOM) algorithm [14] and Latent Semantic Indexing (LSI) provided similarity clusters for visualising patterns across the artefacts. Furthermore a tool for developing a 'soft ontology layer' was developed, allowing simple non-hierarchical categorisation by a community of users. The ontological models can be compared to the ontological spectrum in figure 1. Firstly, the Soft Ontology Layer provides a non-hierarchical categorisation an artefacts by a group of users. The use of LSI to develop semantic spaces also provides a more descriptive method as found towards the right hand side of the spectrum. By combining these two approaches the Nordic forum attempts to overcome the short comings that the use of one method alone would give.

4.2 FORUM CELEBRATING TECHNOLOGY INNOVATIONS IN SOUTH CENTRAL ENGLAND

South Central England has a long tradition of technological innovation. The Knowledge Media Institute in association with the Bletchley Park Museum developed a forum celebrating this tradition. Bletchley Park was the home of the British government's Code and Cypher School during the Second World War and was the centre responsible for analyzing and decrypting Axis signals. The forum focused on supporting a group of about 35 Bletchley Park tour guides in their investigations and promotional works for the museum. The work done in developing the Bletchley Park forum is described in detail in, [15], [16], [4] and [17].

4.2.1 DOMAIN CONTENT

The domain content for the forum comes from the Bletchley Park Museum and relates to wartime code breaking and early technological innovation. The forum aims at exploiting the pervasiveness of stories as an essential part of how we think. Within this context the forum contains about 200 annotated interviews with World War II code breakers and narratives authored by the Museum guides themselves.

4.2.2 APPLICATION OF KNOWLEDGE SUPPORT (STORY FOUNTAIN)

The forum consisted of an annotation and exploration tool called 'Story fountain'. Two assumptions were made in providing intelligent support through the Story Fountain tool. Firstly, exploration of the domain would be question driven rather than unguided and secondly the questions being explored would be open-ended and evaluative rather than solely

A Carta Marina was digitized version as part of the CIPHER project and can be found at http://cipher.uiah.fi/forum/materials/ carta_marina/annotations/

factual. To support these activities different methods of organizing and traversing the story archive were required – these methods can be matched to the task and characteristics of the user. The exploration functionality of *Story Fountain* was supported through the extension of a standard ontology called the Conceptual Reference Model² (CRM) [18]. The CRM was developed by the *International Committee for Documentation* of the *International Council of Museums* (ICOM-CIDOC). The CRM provides a formal high level set of concepts, properties and relationships which can be applied to the broadest range of cultural heritage domains. Through its implementation the forum members were supplied with increased functionality, highlighting the capabilities of formal representations. The CRM was also accepted by ISO and became the base of the draft of an international standard ISO/CD 21127.

4.2.3 SUPPORT FOR DYNAMIC NARRATIVE AUTHORING

The *Story Fountain* integrates a number of existing or emerging authoring tools under the term of Dynamic Narrative Authoring (DNA). The tools support the following functionality:

Time reasoning support: This functionality is supported through complex reasoning by the forum's inference engine. DNA aims at assisting the user in developing time consistent narrative or carry out further research into the correctness of existing narrative.

Annotating resources: Both the RAT (Resource Annotation Tool) and PAT (Picture Annotation Tool) tools are used during the interpretation process allowing the author to organize their narratives and store them in the database. Both tools are described fully in [8].

Support for conceptual graphs: allowed the user to create a narrative using the DNAT tool describe in section 4.4 and import the narrative and corresponding ontology into the *Story Fountain* tool for public exploration and publishing.

The *Story Fountain* tool provided advanced exploration techniques through the highlighting of story concepts contained with each narrative and the relationship between these concepts. The functionality displayed in figure 3 was provided coupled with the ability to map events across stories.

4.3 SHARED HERITAGE OF CENTRAL EUROPE

The Central European forum was developed by Czech Technical University in Prague in association with RiSKommunal in Austria. As it is quiet a large and culturally diverse region four communities participated in capturing the region's shared heritage. They are the *Historical Guild of Schwarzenberg*, the regional Municipalities, the Historians of Southern Bohemia and Art history students from Charles University in Prague and Palacky University in Olomouc. The four communities were supported in the collection, annotation and retrieval of web resources by innovative tools and techniques as described in, [19], [20] [21] and [22].

4.3.1 DOMAIN CONTENT

The domain content is broken down by the interests of the different communities involved in contributing to the forum.

- *Regional municipalities* consist of chiefly of users of the RIS³ Municipal Information System situated in the Czech Republic and Austria.
- *Historical guild of Schwarzenberg* is a historical community that was established at the beginning of the last century by tenants and former employees of the Schwarzenberg family.
- Art history students consist of a group of undergraduate students in Art history.
- *Historical monument protection agency* is provided by the National Monument Institute, a governmental agency whose task it is to take care of all the national heritage monuments within the Czech Republic.

4.3.2 APPLICATION OF KNOWLEDGE SUPPORT

The Central European forum again took a different approach in supplying knowledge support. The forum consists of four communities each specializing in a separate area of interest. Each community was dealt with separately by the forum designers yet when brought together represented the shared heritage of the region. Both the historical monument protection community and the art history group developed (with the help of the forum developers) hierarchical slot-based ontologies representing their respective domains. The historical monument protection community developed an ontology of chapel facades which described the terminology needed for describing all the basic chapel typologies of the region. Similarly, the art history students, aided by a professional historian from the National Monument Institute developed an

²

For more information on the CIDOC CRM see http://cidoc.ics.forth.gr/

³ http://www.riskommunal.net/

ontology representing a typology of palace facades. Again the ontology was a hierarchical and slot-based one as seen towards the right hand side of the McGuinness spectrum (figure 1). Both communities were supported by the CIPHER knowledge cycle through the use of Apollo for ontology development, RAT-O (Resource Annotation Tool – Out) [22] and PAT-O (Picture Annotation Tool-Out) for annotation and publishing. Following the collaborative development of the domain ontologies, both communities used the RAT-O tool to develop online presentations.

In collaborating with the Knowledge Media Institute, the Czech Technical University implemented a version of the *Story Fountain* tool to aid in the exploration of stories from South-Bohemian – the resulting application was called *South-Bohemian Story Explorer* and provided a means for the historians from the region to develop narratives promoting the area's heritage. Reasoning support provided through the *Story Fountain* tool was exploited by the *South-Bohemian Story Explorer*. The ontology configuration for the forum followed the modularity rule of knowledge modelling which resulted in a layered ontology structure [23]. Again CIDOC's abstract Conceptual Reference Model (CRM) provides the upper ontology. Conceptual modelling follows the principle of moving from the abstract to the specific, hence allowing reuse of the upper ontologies at a later date. However, abstract models may provide a starting point and an excellent resource for semantic interaction between heterogeneous sources, but more specific domain ontologies are required to provide the more definitive concrete concepts in a specific focused domain.

4.3.3 SUPPORT FOR DYNAMIC NARRATIVE AUTHORING

The Czech Technical University developed a software tool titled DNAT or Dynamic Narrative Authoring Tool. The DNAT tool was designed to support users in creating 'knowledge intensive content', such as a CH narrative and a corresponding knowledge base of domain concepts at the same time. [24] The tool allows the development of conceptual graphs, a knowledge formalism built on a human readable format of first order logic (FOL). Conceptual graphs [25] provide a means for representing the principle story concepts and the inter-relations within a narrative. The result is a knowledge intensive narrative with the associated graph represented in a frame based ontology format. The annotated narrative and corresponding ontology can then be imported into the Story Fountain tool for public consumption

5. T.ARC.H.N.A

Tarchna (Towards Archaeological Heritage New Accessibility) is the Ancient Etruscan name of the city of Tarquinia. The appellation is being used in a project to create a 'Virtual Museum', providing new accessibility to Archaeological Heritage. It is intended as a way of providing visitors to the virtual museum with a better interpretation of the exhibits through the use of specific cultural heritage narratives and 3D technologies. The project will do this by developing a comprehensive database of Etruscan cultural heritage artefacts and monuments and a virtual Museum interface to provide a comprehensive tour of the objects kept in European Museums. In part borrowing benefiting from the research carried out on CIPHER the project, it aims to utilise explicit narrative concepts and the descriptive power found in the CIPDOC CRM to provide advanced features for visitors to the virtual Museum.

6. SUMMARY

Four forums were introduced in this paper – each highlighting new and innovative approaches to the use of ontologies in supporting online communities within the cultural heritage sector. Each approach emphasised a need for an awareness of explicit narrative objects and each implemented support for those narrative objects. Through the use of ontologies and narrative concepts each forum provided an engaging experience for the community of interest. As this research continues, many of the concepts highlighted here are been utilised on the TARCHNA project.

ACKNOWLEDGEMENTS

This research was conducted within the EU IST funded CIPHER project, IST-2001-32559.

REFERENCES

GRUBER, T. (1993) - What is an Ontology?

MCGUINNESS, D. L. (2002) - Ontologies Come of Age, in Spinning the Semantic Web, Diete, Editor. The MIT Press.

p. 392.

CIPHER, partners (2004) - D25: Second Trial Public Presentation.

MULHOLLAND, P., et al. (2003) – D16: Support for dynamic narratives.

KILFEATHER, J., et al. (2003) - Cultural Heritage Forum Desgin. in ICHIM. Paris.

KILFEATHER, J., et al. (2003) - An ontological application for archaeological narratives. IEEE Intelligent Systems.

CORNS, A.; KILFEATHER, E. (2003) – Internet cultural portals and the development of an Irish Cultural Heritage Forum. in CAA Enter the Past Conference. Vienna.

PETR MIKŠOVSKÝ; LUBOŠ KRÁL (2003) – D15: Tools for communities of interest – CTU.

MULHOLLAND, P., et al. (2003) - D6: Personalisation. Knowledge Media Institute: Milton Keynes.

DÍAZ-KOMMONEN, L.; KAIPAINEN, M. (2002) – Designing vector-based ontologies: Can technology empower open interpretation of culture heritage objects? in IEEE International workshop on Presenting and Exploring Heritage on the Web (PEH'02) in conjunction with the 13th International Conference and Workshop on Database and Expert Systems Applications. Aix-En-Provence, France.

COLLAO, J. (2003) – Soft Ontologies and Similarity Cluster Tools to Facilitate Exploration and Discovery of Cultural Heritage Resources. in DEXA. 2003. Prague.

KAIPAINEN, M.; DÍAZ-KOMMONEN, L. (2002) – Random Vector Encoding Applied to Cultural Heritage Domain. Media Lab UIAH: Helsinki.

PUTNAM, J. (2001) - Open the box, in Art and Artifact - The Museum as Medium. Thames & Hudson. p. 26.

KOHONEN, T. (1982) – Self-organizing formation of topologically correct feature maps. Biological Cybernetics, 1982. 45 (5969).

COLLINS, T. D.; MULHOLLAND, P. (2002) – A Cultural Heritage Forum Celebrating Technological Innovation at Station X. in 13th International Conference and Workshop on Database and Expert Systems Applications (DEXA 2002). Aix-En-Provence, France.

MULHOLLAND, P., et al. (2002) – CIPHER: Enabling Communities of Interest to Promote Heritage of European Regions, in cultivate interactive.

MULHOLLAND, P.; COLLINS, T. D. (2004) – Using Digital Narratives to Support the Collaborative Learning and Exploration of CH. 2004: p. 5.

CROFTS, N., et al. (2004) - Definition of the CIDOC Conceptual Reference Model. ICOM/CIDOC. p. 94.

WOLFF, A., et al. (2002) – Tools for personalised annotation and organisation of diverse Web resources. in Database and Expert Systems Applications.

KRÁL, L., et al. (2003) – Family of Software Tools for an Internet Portal. 2003, Department of Cybernetics, The Gerstner Laboratory, Prague

CIPHER, PARTNERS (2003) – D5: Presentation of Application Domains.

MATOUSEK, K., et al. (2003) – Resource Annotation and Outline Creation Tool (RAT-0). in 14th International Workshop on Database and Expert Systems Applications.

AUBRECHT, P.; KRAL, L.; MATOUSEK, K. (2004) - Collaborative Discovery - D18. CTU: Prague.

UHLÍR, J., et al. (2004) - DNAT - User's Manual. CTU: Prague.

SOWA, J. (2000) – Knowledge Representation: Logical, Philosophical and Computational Foundations, Course Technology.

FIGURES



Fig. 1 - Ontology Spectrum (after McGuinness).



Fig. 2 - Stages in exploring cultural heritage [8].