John Mc Auley – James Carswell

An Open Approach to Contextualising Heterogeneous Cultural Heritage Datasets

Abstract: This paper introduces the TARCHNA system, an approach to presenting contextualised heritage content. The system uses the CIDOC CRM ontology to consolidate a virtual repository of geographically disparate heritage databases and can present a holistic view of a fragmented heritage. While previous approaches to presenting heritage collections have focused on the browse and search paradigm, the TARCHNA system uses narrative concepts as a means of presenting contextualised heritage artefacts within a broader cultural setting.

Introduction

As with other areas of information management, curators and museum professionals use a variety of approaches and systems to manage their digitised content. The conventional problems that burden the interoperability of heterogeneous datasets are therefore highly significant to the domain of cultural heritage. Supplementary is the difficulty of "context": i.e. presenting heritage artefacts from a broader perspective and within their original context. Antecedent approaches have addressed this problem by focusing on presenting a united view of museum collections. Conversely, we suggest capturing the domain expert's interpretation by way of narrative presentations, and conveying a unified and contextualised portrayal of a cultural heritage domain. Although the TARCHNA project focuses on Etruscan heritage, the approach is considered general enough to be applied across the sphere of cultural heritage.

This paper gives an overview of the TARCHNA system architecture, the tiered components, and reasons behind the approach. An explanation detailing the issues of context is provided, plus the proposed solution of using ontologies to describe collections, cultural and narrative concepts. The paper concludes with a description of how mapping between the CRM ontology and a TARCHNA database was implemented.

TARCHNA System Architecture

Typical of enterprise information systems, the TARCHNA system is divided into a multi-tiered architecture whereby each tier supports a clear division of labour.

Source

The source tier consists of several heterogeneous data-sources, each exhibiting a separate (i.e. dissimilar) database schema, and three ontologies developed upon the CIDOC CRM data standard (Crofts et al. 2005). The first or TARCHNA Domain Ontology provides a common reference model on which to integrate the different database schemas. This integration takes the form of a "Database to Ontology" mapping, whereby the elements defined by the database schema are expressed in terms of the ontology's concepts. Unlike other efforts, such as ARTISTE (Allen / Vaccaro / Presutti 2000), the database mappings are held in the ontology. In this way, the mapping information can be accessed in the same way as instances of the ontology classes – thus removing the need for an external procedure to access mapping files or altering the individual database schemas. Two other ontologies, "Sphere of Knowledge" and "Narrative Ontology", also sit in the source tier of the system; each is discussed in later sections. All three ontologies are represented in the RDF formalism and stored in a sesame RDF store (Broekstra / Kampman / van Harmelen 2002).

Knowledge

The knowledge or second tier is the central constituent of the system. It consists of several authoring tools, which support domain experts to develop narrative content, and the TARCHNA engine, which guides interaction between the TARCHNA Domain Ontology and the individual databases. The TARCHNA engine processes requests expressed as ontological concepts, and converts them into sepa-
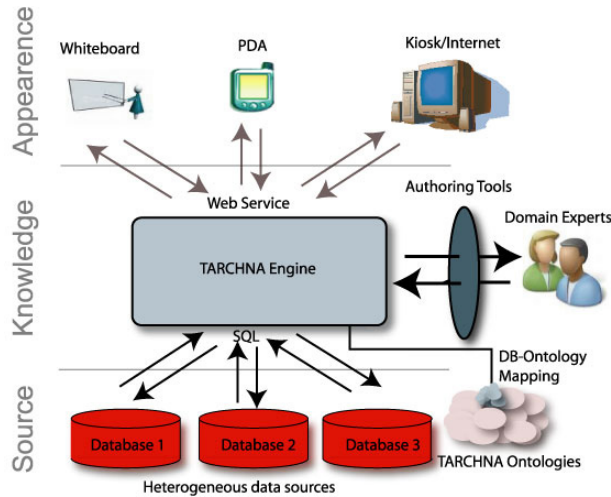
Fig. 1. TARCHNA system architecture.

Fig. 2. Developing a narrative presentation with the TARCHNA authoring tools.

rate SQL statements relevant to each dataset. This process occurs at two different stages: firstly when authors, wishing to write a narrative, search the system for relevant artefacts on which to base their narrative content; and secondly when a request is accepted from the TARCHNA web service, the engine retrieves all narrative content related to a specific artefact, or conversely, retrieves all artefacts related to a specific narrative. The TARCHNA engine and authoring tools were developed in the Java programming language with the Protégé Ontology API (Noy / Fergerson / Musen 2000; Gennari et al. 2003).

Appearance

The third and final tier acts as the disseminating component of the system, and distributes a data (i.e. artefacts) enhanced narrative to several multimedia devices known as virtual wings (VW). The XML web service technology is used as means of interfacing the appearance layer with the virtual wings. It was felt that a service orientated and platform neutral architecture (SOA) supports a clear demarcation between the internal workings of a virtual wing and the overall data model of the system. In this way, new virtual wings may be added without a reliance on proprietary software or adjustments to the system architecture. Virtual wings are currently conceived as operating within three possible spheres:

- Firstly, as contextualised panoramic images. This innovative approach supports a comprehensive way of integrating conceptual models, such as the TA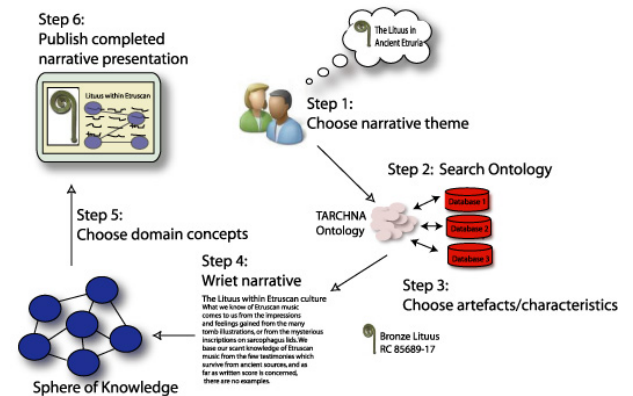RCHNA Domain & Narrative Ontologies, into panoramic images. The approach specifies semantic hotspots or trigger points whereby a visitor can query the image and receive information from the semantic model. It offers a new paradigm for accessing and interacting with semantically contextualised multimedia (Mazzoleni et al. 2006).

- Secondly, as handheld interactive tools. As both GPS and PDA technologies evolve into lightweight and economical location aware handheld devices, it has become increasingly possible to develop high bandwidth GPS applications for mobile devices. This VW is thought of as offering visitors a unique opportunity to explore ancient sites through real-time GPS based digital narrative, and thought of as similar to a personal guide.

- Thirdly, as a customised virtual museum. This option consists of narrative-rich multimedia-based applications operating within a museum space.

TARCHNA Approach to Narrative

Narrative is proposed as a way of reconciling physical artefacts with their original intention or historic context and, in our case, presenting a holistic impression of Etruscan heritage. The aim is to support a team of domain experts (archaeologists, researchers, etc.) develop narrative presentations, which describe artefacts and their context within Etruscan society. In discussions with several archaeologists, the problem of assigning context to digital artefacts was raised. They suggested that in a cultural heritage setting, an artefact's context can be understood

as a combination of its function and role within a specific society. From this perspective, artefacts are presented as references to physical objects from the underlying datasets, many of which are accompanied by multimedia illustrations, while their context is woven into the narrative text and buttressed with ontology concepts, representing both function and role, from the Sphere of Knowledge (ontology).

Authoring Scenario

The authoring process, illustrated in *Fig. 2*, takes place in the knowledge tier of the system. Domain experts are provided with a personal narrative space, in which they can add, edit, and delete narrative content. The authoring process is divided into a number of steps, each contributing towards a completed narrative presentation.

The following scenario illustrates how a domain expert (*Tom*) contributes narrative content to the TARCHNA system:

Tom, an archaeologist working on a dig in Tarquinia (an ancient city in Italy), wishes to discuss Etruscan musical instruments as a whole but would like to focus on the Lituus[1] as an example of one of the "fruits" of Etruscan culture. He approaches the system with several concepts in mind. Naturally, he wishes to feature the artefact itself, and would further like to discuss its role and function within Etruscan society. In step 1 of the authoring process he is asked to choose the theme of his narrative. The theme denotes the nature or broad idea of the text and, from the systems point of view, illustrates the author's interests. For this example, *Tom* wishes to discuss musical instruments and therefore chooses the theme *Finding* from the list presented in *Fig. 3*.

Each theme is supported by a faceted search interface, which uses concepts from the TARCHNA Domain Ontology to specify the search criteria across all heterogeneous datasets. In this example, *Tom* is presented with the search interface specific to the theme of *Finding (Fig. 4)*. The interface highlights a number of characteristics relevant to archaeological findings, such as provenance, inscription, depiction, etc. *Tom* is generally interested in musical instruments, and therefore chooses this concept as the finding class. He further specifies the shape as Lituus and the material as bronze (as is often the case with this type of musical instrument). When satisfied with the choice of criteria, he submits the search to the system.

At this point, the TARCHNA engine *(Fig. 1)* receives the search criteria and translates the ontology concepts into individual SQL statements relevant to each of the heterogeneous datasets. This process involves querying the TARCHNA Domain Ontology (stored in RDF) for the mapping information applicable to each database. The resulting RDF triples are transformed into SQL statements and each database is queried. The results are then returned to the TARCHNA engine, where they are correlated into a single resource and sent back to the author, as demonstrated in *Fig. 5*.

The results of *Tom's* search are divided under the headings of Direct and Indirect Narrative. While Direct Narrative discusses artefacts with reference to specific database objects, an Indirect Narrative discusses characteristics of artefacts as represented by concepts within the TARCHNA Domain Ontology (e.g. Shape: Lituus, or Material: Bronze). There are several incentives for this approach. Firstly, authors are supported in discussing artefacts from a general perspective, without relying on reference to a specific database object. This could amount to a discussion on Etruscan musical instruments, without explicit artefact references, but with an indication as to the shape of Lituus for example and consequently to any artefact of that type. It is suggested that the approach could draw on a more active participatory role from the reader, as the narrative acts as a gateway to further exploration of, in this case, Etruscan musical instruments. Secondly, the concept of Indirect Narrative supports collections that may be added to the TARCHNA system at a later date. For example, let's consider artefacts with the shape of a Lituus that are discussed by an Indirect Narrative. If a new collection is added and, following the mapping procedure, there are new artefacts of shape Lituus present, those artefacts are immediately associated with that Indirect Narrative. Thirdly, an author wishing to contribute to the system's content is not discouraged from doing so by the absence of a particular artefact, and is instead proffered with the

---

[1] The Lituus was both a crooked staff, usually held by influential individuals, and an L-shaped wind instrument. Although it functioned as a musical instrument, its role was often during religious rituals.
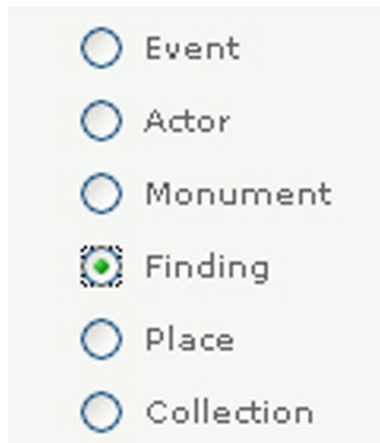
Fig. 3. Step 1: choose from a list of several themes on which the author wishes to base their narrative presentation.

opportunity to contribute, albeit from a more general perspective.

Returning to the example, it can be seen from *Fig. 5* that, in this instance, *Tom's* search yields a reference to a bronze Lituus with the uid *RC 85689-17. Tom* decides therefore to concentrate on a Direct Narrative discussing the value of the Lituus within Etruscan culture. He chooses the Lituus reference (as illustrated in *Fig. 5*) and proceeds directly to writing his text. Having completed his text, *Tom* chooses the terms from the Sphere of Knowledge, which best describes his narrative content. Again, the Sphere of Knowledge is a less formal ontology illustrating the broader aspects of Etruscan culture. Each term,

Fig. 4. Step 2: the author is presented with a faceted search interface specifying a number of characteristics related to Etruscan findings. The interface is made up of concepts from the underlying Domain Ontology.

Fig. 5. Step 3: the results of a search for describing the wind instrument Lituus. Indirect Narrative allows the author to write about the characteristics of the Lituus without reference to an actual artefact; conversely, Direct Narrative supports the author.

from *Art and Artefacts* to *Environment and Landscape,* represents the top level of the ontology. By clicking on a term, the author expands the ontology tree and a more specialised branch of the hierarchy is displayed. In this example, *Tom* is discussing a particular type of musical instrument, therefore the term *Customs* is chosen, followed by the more specialised term of *Music and musicians* depicting the artefact's function. However, a Lituus had a different role in Etruscan society, it was often used during votive offerings and other religious rituals, and as a result *Tom* expands the term of *Religion* choosing both *Rituals in a sacred context* and *Offering (Fig. 6).*

When satisfied, *Tom* saves the finished product to the TARCHNA system. The narrative is comprised of title, text, a direct relation to the artefact Lituus, and a reference to the function, *Music and musicians,* and the role *Rituals in sacred context* and *Offering.* Once saved it is available for dissemination by the TARCHNA web service.

Conclusions

This paper introduced a unique way of presenting geographically disparate heritage collections. The key advantages of which are listed below:

Fig. 6. Step 5: expanding the ontology branches, *Tom* chooses the terms which best fit his narrative content.

An Open Approach to Interoperability

Separating responsibility between tiers serves a very specific purpose by way of semantic interoperability, and promoting new and exciting ways of accessing cultural heritage information. A key principle behind this approach was that the system relies on no single data model, therefore databases can be added or removed with the minimal of effort. In this way, amendments to a data source do not impact the underlying semantic structure, and through the TARCHNA web service, developers are encouraged to invent new ways of exploring the narrative content.

A Collected View of Distributed Heritage

Heritage collections are often distributed across several, geographically remote, museum databases. By separating functionality between tiers and mapping collections into a single umbrella ontology, the TARCHNA system presents a collected view of a distributed heritage. While users are supported searching multiple datasets, data replication is avoided, and cultural institutions retain tutelage over digitised collections.

Enhanced Data Dissemination with Contextualised Narrative Content

Heritage professionals have acknowledged the importance of "context" when presenting artefacts to the general public. By foregrounding artefacts within a narrative backdrop, it is proposed that objects are considered from a broader contextual perspective.

Acknowledgements

References

Allen / Vaccaro / Presutti 2000
  P. Allen / R. Vaccaro / G. Presutti, ARTISTE: An integrated Art Analysis and Navigation Environment. Cultivate Interactive 1, 2000. http://www.cultivate-int. org/issue1/artiste/ [31 Dec 2007].

Broekstra / Kampman / van Harmelen 2002
  J. Broekstra / A. Kampman / F. van Harmelen, Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema. In: I. Horrocks / J. A. Hendler (eds.), The Semantic Web. Proceedings of the 1st International Semantic Web Conference, Sardinia, Italy, June 9–12, 2002 (Berlin 2002) 54–68.

Crofts et al. 2005
  N. Crofts / M. Doerr / T. Gill / S. Stead / M. Stiff, Definition of the CIDOC Conceptual Reference Model 4.2. Heraklion, June 2005. http://cidoc.ics.forth.gr/docs/ cidoc_crm_version_4.2.pdf [11 Sep 2008].

Gennari et al. 2003
  J. Gennari / M. A. Musen / R. W. Fergerson / W. E. Grosso / M. Crub / H. Eriksson / N. F. Noy / S. W. Tu, The evo-

lution of Protege: an environment for knowledge-based systems development. International Journal of Human Computer Studies 58,1, 2003, 89–123.

MAZZOLENI et al. 2006

P. MAZZOLENI / S. VALTOLINA / S. FRANZONI / P. MUSSIO / E. BERTINTO, Towards a contextualized access to the cultural heritage world using 360 Panoramic Images. Software Engineering and Knowledge Engineering (San Francisco 2006).

NOY / FERGERSON / MUSEN 2000

N. F. NOY / R. W. FERGERSON / M. A. MUSEN, The Knowledge Model of Protege-2000: Combining Interoperability and Flexibility. In: R. DIENG / O. CORBY (eds.), Knowledge Acquisition, Modeling and Management. Proceedings of the 12th International Conference, EKAW 2000, Juan-les-Pins, France, October 2–6, 2000. (Berlin, Heidelberg 2000) 17–32.

John Mc Auley
James Carswell

Dublin Institute of Technology
Digital Media Centre
Aungier st
D2 Dublin
Ireland
john.mcauley@dit.ie