Topic Maps as Application Data Model for Subject-centric Applications

Gerhard E. Weber, Ralf Eilbracht, and Stefan Kesberg

Hoelle & Huettner AG, Tuebingen, Germany[1]

Abstract. Today most applications operate on backends with application specific data models. In contrast to this, we suggest modelling application specific information structures at the level of content, and not at the level of the data model. We demonstrate our approach with a publicly accessible web application. Based on a domain ontology, and a set of knowledge models all content for the example application was mapped into a Topic Map. A Topic Maps web frontend renders interface structures, and knowledge-oriented access paths to the highly networked information space of its backend, and also provides relational tables as if it was based on an application specific data model. This approach provides flexible storage layers for applications, and allows using a single data model for different applications. Moreover, applying subject-orientation from high level ontological concepts down to the data level of property values changes accessing content from navigating data-oriented application specific frontends to navigating knowledge-maps.

Keywords: ontology-based application, subject-centric computing, Topic Maps-browser, Topic Maps-frontend, ecotoxicological ontology, Topic Maps-based application.

1. Introduction

Today, most application systems are based on backends with application specific data models, and user interfaces are closely coupled to the relational models of their backends. However, requirements are not static, and very often data models as well as frontends need to be adjusted at high costs. Another disadvantage of application specific data models is the high cost for integrating resources in backends with different models, which is increased by the usually lacking information on the semantics of their constructs. Therefore, we looked for an alternative to building application systems on top of backends with application specific data models and dubious semantics.

---

[1] Current address: Nexxor GmbH, Stuttgart, Germany; {gerhard.weber, ralf.eilbracht, stefan.kesberg}@nexxor.de

Maicher, L.; Garshol, L. M. (eds.): *Subject-centric computing. Fourth International Conference on Topic Maps Research and Applications, TMRA 2008, Leipzig, Germany, October 16-17, 2008, Revised Selected Papers.* (Leipziger Beiträge zur Informatik: XII) - ISBN 978-3-941152-05-2

Another option for building application systems would be using a single, and preferably a standardized data model for different applications. Such a data model should enable the attachment of semantics to its constructs. In addition, a standardized processing logic for integrating content in distinct database instances of the data model would provide another benefit with respect to integration requirements. While the Topic Maps data model [6] is aimed at modelling knowledge and connecting encoded knowledge to resources, it is rich enough to be used for modelling knowledge-oriented as well as data-oriented content. It does provide mechanisms for attaching semantics to content and it also defines the sought after processing logic for merging. Therefore, we explored the use of the TMDM [6] as application data model, and demonstrate this approach with a publicly accessible web application developed for the German Federal Environment Agency [11].

2. Using Topic Maps as Application Data Model

The TMDM [6] is a graph-based data model defining a small number of information item types. Out of a total of seven Topic Maps constructs, three fundamental constructs are topics, associations, and occurrences.

A topic is defined as "a symbol within a topic map, which represents a subject about which assertions are to be stated" [6]. Topics may represent "anything whatsoever". That is, topics may represent high level universals like *method*, *object*, *process*, *kind of property* or individuals like a particular sample, a particular property of a particular sample, or even the value of a particular property of a particular sample.

An association is a typed "representation of a relationship between one or more subjects" [6]. The standard defines subject identifiers for the type-instance relationship as well as for the supertype-subtype relationship. With the supertype-subtype relationship hierarchical relationships like the specialisation relationship can be modelled based on standardized subject identifiers. The supertype-subtype relationship is well suited for modelling knowledge-structures such as taxonomies.

An occurrence represents a typed "relationship between a subject and an information resource" [6]. For a topic of type *mean geometric property value*, an occurrence of type *repeatability standard deviation* may be defined, and the value of this occurrence may have a particular data type such as *xs:double*. By providing data types at the level of occurrences, Topic Maps offer an internal bridge over the gap between knowledge-oriented and data-oriented content.

Another Topic Maps construct internally bridging that gap are variant names, the values of which may also have a data type.

The standardized merging process of Topic Maps, supports a stepwise and layered approach for the development of applications. In a first phase a topic map containing the application ontology is created. In a second phase topic maps containing the domain knowledge models are created. In a third phase topic maps with data-oriented content are created. For the productive application all topic maps at the ontology, the knowledge model and the data-oriented level may be merged into a single topic map.

3. An Example for an Application based on Topic Maps

3.1 Background

In the European Union, so far no methodological recommendations for the assessment of the ecotoxicity of waste have been provided. Therefore, the German Federal Environment Agency (FEA) coordinated a European ring test for the evaluation of a battery of methods for assessing the ecotoxicity of wastes and waste eluates described as hazard criterion H14 in the European waste list [[, 9]. The results of this ring test are intended to support the drafting of binding European recommendations for the assessment of the ecotoxicity of waste.

Given the scientific value of the ring test results, and their intended use as expert inputs to strengthen the basis for European environmental policy, all results were to be published online. However, due to the high complexity of the ecotoxicological domain, and the project and data structures, a web-frontend based on an application specific relational data model seemed inappropriate to assure easy accessibility of the expert results for non-experts. Instead, FEA opted for an approach of an entirely Topic Maps-based backend for all information about, and for all results of the ring test. Easy accessibility of the complex data was a prime requirement intended to improve not only expert usage of the results but also the political impact achievable with the outcome of the project[2].

The ring test focused on three waste substrates which were evaluated by laboratories all over Europe. All in all 67 laboratories participated in the ring test, and 17 different ecotoxicological methods using 16 different biological species were applied. Including the reference substances tested, close to 200 different properties were assessed. The numerical results were subject to statistical

2 The H14-Navigator is a commercial customer specific web application created by Hoelle & Huettner AG for the German Federal Environment Agency; to be accessed at `http://EcotoxWasteRingtest.uba.de/h14`.

analysis, which provided the basis for the evaluations of the ecotoxicological methods employed and for the methodological recommendations for assessing ecotoxicity of waste.

3.2 Application Specific Ontology

As an interdisciplinary science, ecotoxicology draws from concepts in a variety of domains such as ecology, biology, chemistry and toxicology. Although, chemical and physical properties of the ring test samples have been excluded from the online publication so far, the core of the ontology was shaped in a way that would allow the mapping of both additional property domains. Therefore, an ontology of physical, chemical and biological properties published by Dybkaer [4] was adapted to the needs of the ring test project and its results.

We aimed for a realistic ontological approach rooted in high level universals such as *object*, *process*, *method*, *property*. For linking these high level universals to the low level universals characterizing the entities dealt with at the laboratory level, we used the superclass-subclass relation.

The adapted ontology on property [4], as well as the project ontology and the ontological components for the biological entities to be modelled was mapped into Topic Maps constructs.

3.3 Integrated Knowledge Models

The sole examination principle used by the ecotoxicological methods studied in the ring test was the response of living organisms exposed to samples of the waste substrates. Considering that taxonomic relations between the species used in the ring test are highly relevant knowledge structures for the test results, a phylogenetic tree was mapped into the Topic Map. It comprised the 16 biological species used in the ring test, and can be considered as one of the core knowledge structures of the ecotoxicological domain. The phylogenetic tree also served for structuring the specialisation hierarchies of the top universals method, process, and property.

3.4 Topic Maps Benefits for Interface Structure and Functionality

All content including results at five different levels from the laboratory level to the level of the ring test recommendations, as well as project structures, the ontology, and all knowledge models were mapped into a single topic map. Due to the use of association-type names scoped by role-types, and their rendering scoped with the role-type played by the focus topic [8], the well known good readability of binary relations between topics is achieved (Fig. 1).

In the so-called topic view the information displayed for the focused topic is supplemented by visualisations of the relevant knowledge models in order to anchor the subject in the knowledge domain. E.g. the particular Daphnia test displayed in Figure 1 is complemented by the relevant process hierarchy. Thus, even users less familiar with ecotoxicology may recognize that Daphnia tests are ecotoxicological examination processes. Moreover, the process hierarchy offers additional access paths to all instances of its members. Clicking at the high level process type *aquatic ecotox. test* displayed in the hierarchy shown in Fig. 1 would retrieve all tests of this type.

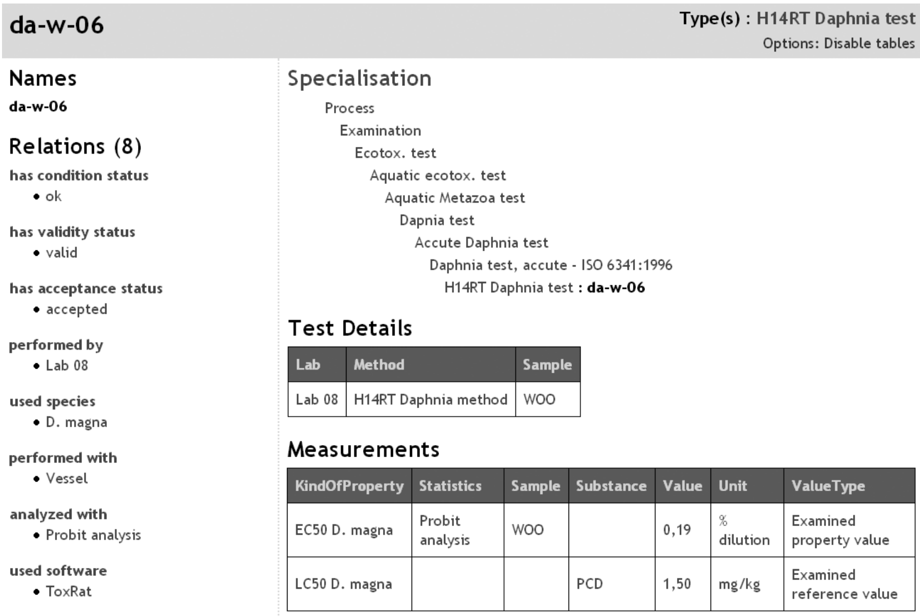
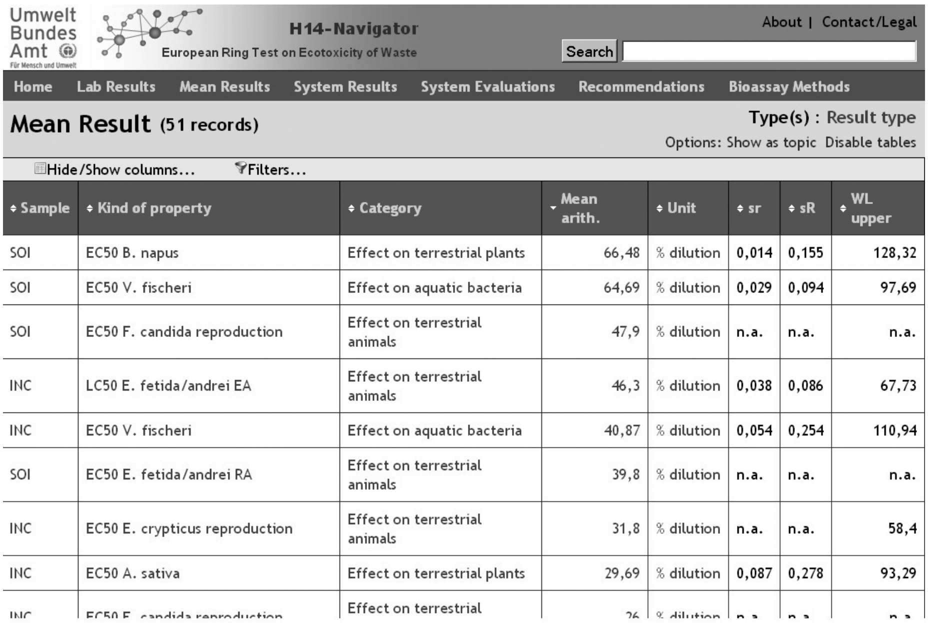


Fig.1: Screenshot detail of a topic view rendered by the H14-Navigator presenting results of a single Daphnia test; the specialisation hierarchy anchors the process in the ecotoxicological domain.

Due to the complexity of the ecotoxicological properties, as well as the ring test structure, views restricted to a single topic, its associations, occurrences, and/or instances are not appropriate to cover all requirements of the application. Relational views are required to visualize results of the ecotoxicological methods used to analyse the three samples. The application specific frontend was therefore enabled to render tables, which provide the same structures as a frontend using an application specific data model would (Fig. 2). Most cells in these tables do not just contain data of a particular data type but topics representing their subjects of discourse. Users may therefore find more information about any topic displayed, by visualising the linked topic views. E.g. each arithmetic mean value of the sample properties depicted in the table shown in Fig. 2. links to its topic view which visualises its name, published subject indicator, occurrences, associations and associated topics. Thus, the highly networked character of the information related to a mean property value is accessible via the value itself rendered in the user interface.



Sample	Kind of property	Category	Mean arith.	Unit	sr	sR	WL upper
SOI	EC50 B. napus	Effect on terrestrial plants	66,48	% dilution	0,014	0,155	128,32
SOI	EC50 V. fischeri	Effect on aquatic bacteria	64,69	% dilution	0,029	0,094	97,69
SOI	EC50 F. candida reproduction	Effect on terrestrial animals	47,9	% dilution	n.a.	n.a.	n.a.
INC	LC50 E. fetida/andrei EA	Effect on terrestrial animals	46,3	% dilution	0,038	0,086	67,73
INC	EC50 V. fischeri	Effect on aquatic bacteria	40,87	% dilution	0,054	0,254	110,94
SOI	EC50 E. fetida/andrei RA	Effect on terrestrial animals	39,8	% dilution	n.a.	n.a.	n.a.
INC	EC50 E. crypticus reproduction	Effect on terrestrial animals	31,8	% dilution	n.a.	n.a.	58,4
INC	EC50 A. sativa	Effect on terrestrial plants	29,69	% dilution	0,087	0,278	93,29
INC	EC50 F. candida reproduction	Effect on terrestrial	26	% dilution	n.a.	n.a.	n.a.

Fig. 2: Screenshot of the web application H14-Navigator rendering a table with results of the H14 Ring Test; most cells of the table represent topics.

As a further contrast to data-oriented applications, the "data" represented in this knowledge-oriented application are equipped with identities, since published subject identifiers are defined for all topics. The well defined identities of the information items prepare the ground for future integration of content.

Both visualisation patterns – the topic view (cf. Fig. 1), as well as the table view (cf. Fig. 2) – benefit from the integrated knowledge-models. Topic views are complemented with the relevant knowledge models by a visualisation of complete branches linking specialized concepts with more general concepts. In table views a particular level of a relevant hierarchical knowledge model may be visualized in order to support relating the displayed records in a knowledge domain. E.g. for each record displayed in Fig. 2. a *category* is computed by a recursive inference rule retrieving a particular level in the knowledge model for *kind of property*. The first record depicted in Fig. 2. is thus categorized as quantifying an *effect on terrestrial plants*. In addition, the interface object providing this information represents a topic, and thus offers an access path to either a table of all instances of its kind or its topic view.

4. Discussion

We suggested using the Topic Maps data model [6] as application data model, and demonstrated the approach with a publicly accessible web application [11]. Our method results in benefits at the user interface level, which are due to the differences between a conventional data-oriented approach and the knowledge- or subject-orientation of our approach. Whereas a conventional application would provide a number of data-oriented static tables for accessing content, our approach provides a multitude of access and navigation paths based on a domain ontology, and a number of domain knowledge models. In addition, it also provides relational views on content, analogous to the conventional table views. However, due to the consequent subject-orientation of our approach, most objects rendered in the tables are topics, and not just strings or numbers.

One of the core components of the ontology for our ecotoxicological example is the ontology on physical, chemical and biological properties by Dybkaer [4], which we modified according to the application's requirements and mapped into Topic Maps constructs. Given the effort required for developing an ontology well grounded in a scientific field, it seems worthwhile to tap the wealth of open biomedical ontologies such as described by Smith et al. [10] for the development of knowledge-oriented applications in this domain. Moreover, for publicly visible applications the support of integrative access is close to becoming a requirement. Applications using open topic-mapped ontologies in combination with the advantages Topic Maps offer in terms of integration might therefore offer a promising field for innovative developments.

Our approach of using the TMDM as application data model was first advocated by Ahmed [1] who rightly claimed that the TMDM matches the decomposition of application design into a set of interacting objects. He further stated that using Topic Maps as application data model, would allow modifications of application model structures simply by altering the data which provides the application schema, thus removing the need to re-compile or re-populate database tables. In short, this comes down to adjusting an application model by changing content of its backend but not structures of its data model and its backend. As a further aspect of considerable advantage Ahmed [1] stated that a single application programming interface would enable accessing the data of any such application.

Although Ahmed [2] elaborated on these ideas, so far Topic Maps has not played a very prominent role as application data model, which might to some degree be due to the slow progress of the standardization process. However, with stable standards for the data model, and the XML syntax [7], with a functional query language [5], and the ISO standardization process for a host of other Topic Maps related standards well under way[3], the time for speeding up the exploration of the full potential of Topic Maps technology seems right.

5. Conclusions

We presented an approach for an application architecture which is doing away with application specific data models, as well as with the distinction between data-oriented, and knowledge-oriented content. Applications based on the Topic Maps data model hold the potential for taking subject-centrism into the realm of data, and thus for anchoring subject-oriented computing at the data level. For applications based on this approach, accessing content changes from navigating data-oriented frontends to navigating knowledge-maps.

Acknowledgements

We would like to thank Heidrun Moser of the German Federal Environmental Agency for triggering work on the H14-Navigator, and Jörg Römbke of ECT GmbH for providing ring test contents, and support in ecotoxicological issues.

References

[1] Ahmed, K.: Topic Maps - A Practical Introduction With Case Studies. XML Europe 2002 (2002)

[2] Ahmed, K.: Topic Maps for (Open) Source Developers. XTech 2005 (2005)

[3] see: `http://www.itscj.ipsj.or.jp/sc34/open/1025.htm`

[3] Becker, R., Donnevert, G., and Römbke, J.: Biologische Testverfahren zur ökotoxikologischen Charakterisierung von Abfällen - Abschlußbericht. Umweltbundesamt (2008). Available at: `http://www.umweltdaten.de/publikationen/fpdf-l/3415.pdf`

[4] Dybkaer, R.: An ontology on property for physical, chemical, and biological systems. APMIS Suppl 117, 1-210 (2004).

[5] Garshol, L. M.: tolog – A Topic Maps Query Language. TMRA05 International Workshop on Topic Map Research and Applications, L. Maicher and J. Park, eds., pp. 183-196. Springer Berlin, (2006)

[6] ISO/IEC 13250-2: Information Technology - Topic Maps - Part 2 - Data Model. International Organization for Standardization (2006)

[7] ISO/IEC 13250-3, Information Technology - Topic Maps - Part 3: XML Syntax. International Organization for Standardization, (2007).

[8] Pepper, S.: The TAO of topic maps: finding the way in the age of infoglut. XML Europe 2000, (2000). Available at: `http://www.gca.org/papers/xmleurope2000/pdf/s11-01.pdf`

[9] Römbke, J., Becker, R. and Moser, H. (eds.): Ecotoxicological characterisation of waste - Results and experiences from a European ring test. Springer, New York (in press).

[10] Smith, B., Ashburner, M., Rosse, C., Bard, J., William, B., Ceusters, W., Goldberg, L. J., Eilbeck, K., Ireland, A., Mungal, C. J., The OBI Consortium, Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R. H., Shah, N., Whetzel, P., Lewis, S.: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology 25, 1251-1255 (2007).

[11] Weber, G. E., Eilbracht, R. and Kesberg, S.: H14-Navigator uses Topic Maps as application data model. In: Ecotoxicological characterisation of waste - Results and experiences from a European ring test, J. Römbke, R. Becker, and H. Moser (eds.), Springer, New York (in press).