Situated Support for Choice of Representation

Sari E. Hakkarainen
sari@idi.ntnu.no

Anders Kofod-Petersen
anderpe@idi.ntnu.no

Carlos Buil Aranda
builaran@stud.ntnu.no
Department of Computer and Information Science
Norwegian University of Science and Technology

Abstract

As more and more companies are augmenting their data to include semantics it is imperative that the choices made when choosing the modelling language are well founded in knowledge about the language and the domain in question. This work demonstrates how the Semiotic Quality Framework can facilitate the choice of the most suited language for a real world application. Computational and situated features are introduced as an extension to the framework.

1 Introduction

The IT industry is currently changing focus from providing storage, processing and network services to providing knowledge intensive information and services to large numbers of customers. The diversity and multitude of resources and applications on the Web places elaborate requirements on methods and tools for efficient generation, manipulation and compositional usage of information and services. Metadata, ontology/domain model and semantic enrichment can bridge the heterogeneity and facilitate the efficient usage of information assets on the semantic Web [1]. However, a formal, standardised representation of signs and meaning is required [15] for supporting ontologies, i.e. explicit and shared conceptualisations [4] of the domain.

Several general-purpose models for description of web-resources have emerged, where the intention is to facilitate the search, aggregation, filtering, selection, reasoning, and presentation of information assets on and for the (semantic) Web. However, the number of languages and models is large, as is the number of types of prospective applications. Applications can be categorised according to the kind of domain they address (medical, commerce, education, library, oil drilling, etc), the kind of application they target (knowledge management, process monitoring, archival, etc.) or the kind of modelling environment they are supposed to fit in (taxonomies, data flows, data models, process models, goal models, etc.). The span for each of these categories is seemingly endless.

The objective is to develop support for the choice of appropriate Web-based knowledge representation formalism. We evaluate existing representations in general, using an existing semiotic quality framework for conceptual models. We propose computational and situated features as extensions to the semiotic quality framework.

The approach proposed here is further described and exemplified in [6], where a general and situated evaluation of Semantic Web languages are documented.

2 The Semiotic Quality Framework

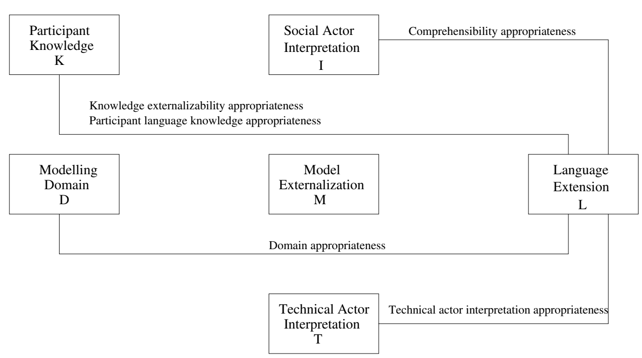

Figure 1. Quality factors in SQF [8]

In order to evaluate the Web representation languages, the *Semiotic Quality Framework* (SQF) [8], [13], a model quality framework consisting of five semiotic factors of quality modelling languages is chosen. The main concepts and their relationships are shown in Figure 1. The framework has three main characteristics that make it well-suited

Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)
0-7695-2415-X/05 $20.00 © 2005 **IEEE**

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 16, 2008 at 10:07 from IEEE Xplore. Restrictions apply.

as an evaluation instrument 1) it distinguish between goals and means separating what to achieve from how, 2) it is closely related to linguistics and semiotic concepts, and 3) it is based on a constructivist world-view, the framework recognizes that models are build from interaction between the designer and the user. The main model of the semiotic quality framework is as follows.

A - **Audience** refer to the individual, A_i, organisational, A_s, and technical actors, A_t who relate to the model. This includes both human participants and artificial actors.

K - **Participant knowledge** is the explicit knowledge that is relevant for the audience *A*. This is the combined knowledge of all participants in the project.

L - **Language extension** is what can be represented according to the graphical symbols, vocabulary and syntax of the language; the set of all statements that may be informal L_i, semi-formal L_s, or formal L_f.

M - **Model externalization** is the set of all statements in an actor's model of a part of a perceived reality written in a language *L*.

I - **Social actor interpretation** is the set of all statements which the externalised model consists of, as perceived by the social audience A_i and A_s.

T - **Technical actor interpretation** is all the statements in the conceptual model *L* as they are interpreted by the technical audience A_t.

D - **Modelling domain** is the set of all statements that can be stated about a particular situation.

The framework evaluates the physical, empirical, syntactic, semantic, pragmatic, perceived semantics, social and knowledge quality; it evaluates the quality of conceptual models, modelling environments, and modelling languages. This work focuses on the evaluation of the Web representations as modelling languages.

2.1 Adapted appropriateness of languages

The Semiotic Quality Framework consists of five quality factors, called appropriateness namely, Domain Appropriateness (*DA*), Participant Knowledge Appropriateness (*PAK*), Knowledge Externalizability Appropriateness (*KEA*), Comprehensibility Appropriateness (*CA*), and Technical Actor Interpretation Appropriateness (*TAIA*). Here we modify the *DA* as in [14], as follows.

DA covers seven perspectives for languages: 1) Structural Perspective refers to the static structure, classes and properties, 2) Functional Perspective refers to the processes, activities, and transformations, 3) Behavioural Perspective refers to the states and transitions between them, 4) Rule Perspective refers to the rules for certain processes, activities, and entities, 5) Object Perspective refers to the resources, processes and classes, 6) Communication Perspective refers to the language actions, meaning and agreements, and 7) Actor and Role Perspective refers to the actor, role, society and organisation.

With the modification of the *DA* we acquire the elements needed for analysing the most practical features of the languages. With the *PKA* we measure the knowledge of the user. With the *KEA* we analyse if the language provides enough elements to represent the domain model specified. With *CA* we analyse if the language is consistent enough and provides clear elements for modelling the domain, and with *TAIA* we analyse if the language provides enough features for allowing automatic reasoning, the key concept in our investigation. The quality factors will be further developed in the next section.

2.2 Selection criteria for quality factors

For the quality of conceptual modelling languages Sindre [12] identifies criteria for the constructs of the language and how these constructs are presented visually. Four main groups of sub-criteria are identified: perceptibility, expressive power, expressive economy, method tools and potential. Seltveit [11] adds the criteria of reducibility, meaning the features provided by the model to handle large and complex models. We propose selection criteria and functions for appropriateness weights.

Let $\mathcal{CF}$ be an evaluation framework such that $\mathcal{CF}$ has a fixed set $\mathcal{A}$ of appropriateness categories *a*, where $\mathcal{A} = \{a_1, a_2, a_3, a_4, a_5\}$ and $a_i \in \mathcal{A}$. Each *a* is a quadruple $< id,\ \ descriptor,\ \ C,\ \ cw >$, where *id* is the name of the category, *descriptor* is a natural language description, *C* is a set of selection criteria *ac*, and *cw* defines a function of *S* that return -1, 1, or 2 as coverage weight, where *S* is a set of satisfied elements *ac* in the selection criteria *C* of each appropriateness category in $\mathcal{A}$. Intuitively, we define a number of selection criteria alongside an associated coverage weight function for each category in the evaluation framework. The appropriateness categories with attached descriptors, selection criteria and coverage weight functions are as follows.

a_1 - **Domain appropriateness (DA)** indicates whether the method guidelines address the problems of eliciting/representing relevant facts of the problem domain. Ideally, $D \backslash L = \emptyset$, i.e. there are no statements in the expected application domain that cannot be expressed in the target language, and one should not be guided to express things that are not in the domain (limited number of constructs). The former criterion means that a_1c_1 - the developer is

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 16, 2008 at 10:07 from IEEE Xplore. Restrictions apply.

guided to make use of high expressive power whereas the latter means that a_1c_2 - there is a limited number of modelling constructs that are generic, composable and flexible in precision. The equation 1 holds for each modelling perspective of a_1p_1 - structural (SP), a_1p_2 - functional (FP), a_1p_3 - behavioural (BP), a_1p_4 - object (OP), a_1p_5 - communication (CP), and a_1p_6 - actor-role (AP) perspective.

$$cw_1(S_1) = \begin{cases} 2 & if \quad a_1c_1 \wedge a_1c_2 \in S_1 \\ 1 & if \quad a_1c_1 \vee a_1c_2 \in S_1 \\ -1 & if \quad S_1 = \emptyset \end{cases} \tag{1}$$

a_2 **- Participant knowledge appropriateness (PKA)** indicates whether the method corresponds to what participant in the modelling activity perceive as a natural way of working. Ideally, $K \cap L \backslash L = \emptyset$, that all the statements in the models of the languages used by the participants are part of their explicit knowledge. Hence a method guideline a_2c_1 - should not promote usage of statements not in a participant's knowledge, a_2c_2 - external representation should be intuitive, and a_2c_3 - non-intuitive representations should be introduced carefully.

$$cw_2(S_2) = \begin{cases} -1 & if \quad |S_2| = 0 \\ 1 & if \quad 0 < |S_2| \leq 1 \\ 2 & if \quad 2 < |S_2| \leq 3 \end{cases} \tag{2}$$

a_3 **- Knowledge externalization appropriateness (KEA)** indicates whether the method assists the participants in externalizing their knowledge. $K \cap L \backslash K = \emptyset$, i.e. there are no statements in the explicit knowledge of the participant in the modelling activity that cannot be expressed in the target language. This appropriateness focuses on how relevant knowledge may be articulated in the language rather than what knowledge is expressed. This implies the partial quality goals of generality, a_3c_1 – the guidance to use the language should be as domain independent as possible, and completeness a_3c_2 – there is guidance for all possible usages of the language.

$$cw_3(S_3) = \begin{cases} 2 & if \quad a_3c_1 \wedge a_1c_2 \in S_3 \\ 1 & if \quad a_3c_1 \vee a_1c_2 \in S_3 \\ -1 & if \quad S_3 = \emptyset \end{cases} \tag{3}$$

a_4 **- Comprehensibility appropriateness (CA)** indicates whether the participants are able to comprehend the method guidelines. Ideally, $L \backslash I = \emptyset$, i.e. all the possible statements of the language are understood by the participants in the modelling effort using the method guidelines. Thus, a_4c_1 - the described modelling constructs are easily distinguished from each other, a_4c_2 - the number of constructs is reasonable or organised in a natural hierarchy, a_4c_3 - proposed use of modelling constructs is uniform for all the statements expressed in the target language, a_4c_4 - the guidance is flexible in the level of detail in the target language, and a_4c_5 - separation of concerns and multiple views is supported.

$$cw_4(S_4) = \begin{cases} -1 & if \quad 0 < |S_4| \leq 1 \\ 1 & if \quad 1 < |S_4| \leq 3 \\ 2 & if \quad 3 < |S_4| \leq 5 \end{cases} \tag{4}$$

a_5 **- Technical actor interpretation appropriateness (TAIA)** indicates whether the method guidelines lend themselves to automated tool support or assist in support for reasoning. Ideally, $T \backslash L = \emptyset$, all possible mechanisms in the technical participants interpretation are supported by the target language. This implies the partial quality goals for automatic reasoning support in the instructions provided for the target language, i.e. a_5c_1 - both formal syntax and semantics are operational and/or logical, a_5c_2 - efficient reasoning support is provided by executability, a_5c_3 - natural language reasoning is supported, and a_5c_4 - information hiding constructs are provided enabling encapsulation and independent components.

$$cw_5(S_5) = \begin{cases} 2 & if \quad a_5c_1 \wedge (a_5c_2 \vee a_5c_3 \vee a_5c_4) \in S_5 \\ 1 & if \quad a_5c_1 \vee a_5c_2 \vee a_5c_3 \vee a_5c_4 \in S_5 \\ -1 & if \quad S_5 = \emptyset \end{cases} \tag{5}$$

The selection criteria for the appropriateness categories above are exhaustive in the categories a_2, and a_4, whereas the set of satisfied criteria S of the remaining categories may also be the empty list. None of the criteria are mutually exclusive. The coverage weight cw is independent of any category-wise prioritisation. Since the intervals are decisive for the coverage weight they can be adjusted depending on preferences of the evaluator. However, when analysing different evaluation occurrences the intervals need to be fixed in comparison, but may be used as dependent variable.

2.3 Weighted quality requirements

Here, we adopt the PORE methodology [9] to prioritise the classification criteria based on company's requirements in order to evaluate the ontology building guidelines in this particular situation. The method has been applied successfully on SQF in [7] for method guideline classification. Hence, the importance weights for each appropriateness category are calculated as follows.

Let $R(CF)$ be a set of weighted requirements such that R has a fixed set $R\mathcal{A}$ of categories ra, where categories in $R\mathcal{A}$ correspond with categories $\mathcal{A}$ of an evaluation framework EF, i.e. $R\mathcal{A} = \mathcal{A}$, and $a \in \mathcal{A}$, $ra \in R\mathcal{A}$. ra is a triple $< id, \quad descriptor, \quad iw >$, where id is the name of the appropriateness requirement category, $descriptor$ is a

IEEE COMPUTER SOCIETY

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 16, 2008 at 10:07 from IEEE Xplore. Restrictions apply.

natural language description of the appropriateness requirement, and iw_{ra} defines a function of I that returns 0, 3, or 5 as importance weight based on priorities and policy of the company, where I is a set of importance judged elements ra in the selection criteria C of each category in $R\mathcal{A}$.

$$iw_{ra}(I) = \begin{cases} 1 & if \quad ra \quad is \quad optional \\ 3 & if \quad ra \quad is \quad recommended \\ 5 & if \quad ra \quad is \quad essential \end{cases} \tag{6}$$

2.4 Situated comparison of languages

When each language in question has been thoroughly analysed through the appropriatenesses a comparison is possible. Each of the total coverage weights Tw_i for each representations i are calculated. The total weights are calculated using Equation 7 and are used as overall feasibility rate for supporting the choice of ontology building guidelines.

$$Tw_i = \sum_{ra \in \mathcal{A}} (cw_{ra} \times iw_{ra}) \tag{7}$$

The weights assigned to the different requirements are compared with the total coverage weights to obtain a ground for selecting the most appropriate language.

3 Conclusions and future work

A situated method for evaluation of representations for Semantic Web applications was proposed extending the ([5], [8]) framework. We argue that the Semiotic Quality Framework (SQF) [8] is well suited for evaluating Semantic Web representations languages. Combined with the use of the numerical values for the weights and adoption of the PORE methodology [9] should produce more explicit evaluation results.

The future objective is to further develop the support for the choice of appropriate Web-based knowledge representation formalism. The way-of-working is to 1) evaluate existing representations in general, using the extended Semiotic Quality Framework for conceptual models, 2) to develop trial ontologies using a common ontology creation tool and the language specifications, and 3) to evaluate the existing representations in an industrial case study. The languages of choice are the Resource Description Framework (RDF) [3], the Web Ontology Language (OWL) [10], and Topic Maps [2]. In the case study in question, the aim is to support the development of an integrated knowledge-based system for directory services by moving from traditional relational data models to semantically richer representations.

4 Acknowledgement

We would like to thank the company Invenio for their cooperation and support in conducting this research.

References

[1] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, 2001.

[2] M. Biezunski, M. Bryan, and S. R. Newcomb, editors. *ISO/IEC 13250:2000 Topic Maps: Information Technology – Document Description and Markup Language*. International Organization for Standardization, 1999.

[3] D. Brickley and R. V. Guha, editors. *RDF Vocabulary Description Language 1.0: RDF Schema*. W3C, 2004.

[4] T. R. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993.

[5] S. E. Hakkarainen, L. Hella, S. M. Tuxen, and G. Sindre. Evaluating the quality of web-based ontology building methods: A framework and a case study. In *6th International Baltic Conference on Databases and Information Systems (Baltic DB&IS 2004)*, Riga, Latvia, June 2004.

[6] S. E. Hakkarainen, A. Kofod-Petersen, and C. B. Aranda. Situated Support for Choice of Representation a Sematic Web Application. Technical report, Departmen of Computer and Information Science, NTNU, 2005.

[7] S. E. Hakkarainen, D. Strasunskas, L. Hella, and S. M. Tuxen. Weighted evaluation of ontology building methods. In *CAiSE 2005 Forum*, June 2005. To Appear.

[8] J. Krogstie and A. Sølvberg. *Information systems engineering - Conceptual modeling in a quality perspective*. Kompendiumforlaget, Trondheim, Norway, 2003.

[9] N. A. Maiden and C. Ncube. Acquiring cots software selection requiremtents. *IEEE Software*, 15(2):46–56, March 1998.

[10] D. L. McGuinness and F. van Harmelen, editors. *OWL Web Ontology Language Overview*. W3C, 2004.

[11] A. H. Seltveit. *Complexity reduction in information systems modelling*. PhD thesis, Norwegian Institute of Technology, 1994.

[12] G. Sindre. *Hicons: a general diagrammatic framework for hierarchical modelling*. PhD thesis, Norwegian Institure of Technology, 1990.

[13] X. Su and L. Ilebrekke. A comparative study of ontology languages and tools. In A. B. Pidduck, J. Mylopoulos, C. C. Woo, and M. T. Ozsu, editors, *Advanced Information System Engineering, 14th International Conference, CAiSE 2002, Toronto, Canada, May 27-31, 2002, Proceedings*, volume 2348 of *Lecture Notes in Computer Science*, pages 761–765. Springer Verlag, May 2002.

[14] X. Su and L. Ilebrekke. *Using a Semitic Framework for a Comparative Study of Ontology Languages and Tools*, chapter 14, pages 278–299. IDEA Group Publishing Group, 2004.

[15] M. Uschold and M. Grüninger. Ontologies: Priciples, methods, and applications. *Knowledge Engineering Review*, 11(2):93–155, 1996.

IEEE COMPUTER SOCIETY

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 16, 2008 at 10:07 from IEEE Xplore. Restrictions apply.