Book Title
Book Editors
IOS Press, 2003

Using context to improve semantic interoperability

Patrick Hoffmann , Lionel Médini and Parisa Ghodous [1]
Lyon Research Center for Images and Intelligent Information Systems
Claude Bernard Lyon 1 University, Villeurbanne, France

Abstract.
This paper presents an approach to enhance interoperability between heterogeneous ontologies. It consists in adapting the ranking of concepts to the final users and their work context. The computations are based on an upper domain ontology, a task hierarchy and a user profile. As prerequisites, OWL ontologies have to be given, and an *articulation ontology* has to be built.

Keywords. context, contextual ranking, semantic resources, semantic similarity

1. Introduction

In an increasing number of organizations, virtual collaboration becomes a reality. More and more collaborative platforms provide means for connecting various models and tools used by different partners of Concurrent Engineering (CE) projects. Computerized data exchanges yet suffer from software incompatibilities, resulting in semantic losses when transmitting high level data. This is particularly true when considering complex data (3D data, simulation data, etc.).

In CE, semantic resources(such as taxonomies, ontologies) are built for specific purposes, and evolve with the projects they are associated to. Semantic interoperability (i.e. interoperability between semantic resources) cannot therefore be achieved by integrating ontologies but by establishing mappings between semantically related concepts from different ontologies. Since semantic interoperability depends on these mappings, it is essential to be capable of evaluating their relevance. As the same mappings are not as relevant for every users and every task, we herein propose an approach to compute a context-based evaluation of such mappings.

Our approach is based on a context model composed of classifications of the organization activity domains and tasks as well as of users' profiles, on OWL ontologies, and on a ranking system. This ranking system receives requests on concepts from the OWL ontologies, made on users' behalf. As results, it returns all semantically related concepts from the OWL ontologies, and ranks them according to the users' contexts. This work is a continuation of Ferreira Da Silva et al.[1] contribution to semantic interoperability with SRILS, a middleware in which the ranking system is intended to be a module.

[1] {patrick.hoffmann, lionel.medini, parisa.ghodous}@liris.cnrs.fr

The paper is organized as follows: in section 2, we examine different propositions to model context, and other related work that use context to compare semantic associations. Section 3 presents our proposal for modeling context and using it to compare semantic associations. Section 4 discusses our methodology to build context models, and its cost. Section 5 concludes the paper.

2. Related work

2.1. Context modeling

The notion of context first appeared as a major principle in CS in [2], where John McCarthy stressed the high significance of the notion, and proposed a formalization of context as a first class object in logic languages. From Artificial Intelligence (AI) fields (Context Modeling, Problem Solving…), the notion is now being used more widely (Information Retrieval, Ubiquitous Computing…). Several definitions of context have been given [3,4,5], that define context as what is relevant to understand a considered event and its implications. More, Dey notes that the final user should to be taken into account as well.

As we are only interested in modeling context and making computations on it to determine the relevance of semantic associations, we think that powerful AI context models are not necessary to our case. Indeed, we are neither primarily interested in preserving consistency nor in making new assumptions by inferences.

2.2. Contextual comparison of semantic associations

Classic comparisons of semantic associations rely on semantic similarity measures or their inverse, semantic distances. Semantic similarity measures are often distance-based (depending on the number of edges that separate the concepts in a graph) [6,7] or information content-based (assuming that some concepts are more informative than others) [8]. Few attempts have been made to contextualize semantic similarity measures. Rodriguez and Egenhofer proposed to match the users' "intention" to increase the relevance of their measure [9]. Roddick et al. underline the necessity of lying on context-dependent similarity measures [10] when computing a global semantic similarity measure.

Aleman-Meza et al. [11], as well as Nejdl and Paiu [12] propose approaches that compare semantic associations in a contextualized manner, not depending on a typical semantic distance. Aleman-Meza et al. consider all entities related to a given entity by a sequence of links, called "paths". Those are ranked depending on both their "universal" and "user-defined" weights. The latter are added when the path traverses a region that the user has set as being of interest, or a source that the user trusts. Nejdl and Paiu propose an authority-based ranking algorithm destined for desktop search. A "context ontology" describes the relations between the files and how they had been obtained (e-mail, url) and used (file properties). Users describe what makes a file important (namely "authoritative"). Files are ranked by applying "authority transfer weights".

We were inspired by the proposition of Rodriguez et al. [9], which leads us to consider users' tasks. As Aleman-Meza et al. [11], we divide semantic resources into regions, and we plan to consider user confidence.

3. Proposal

Our approach consists in a ranking system, a context model and resources gathered by the organization: OWL ontologies, as well as an articulation ontology which contains mappings that connect semantically related concepts of the different ontologies. The ranking system treats requests on a concept from a given OWL ontology, and returns a ranking of semantic relationships between concepts, based on users' domains and tasks.

Our work hypothesis is that individuals' context may be adequately represented by a profile referencing their activity domains and the tasks that they have to perform. This requires that a classification of domains and tasks be built by experts. For our context models, we thus used the OWL[1] sublanguage OWL DL, since it is standard (W3C[2] recommendation), well supported[3], and widely used for writing ontologies. For a ranking system, response time is critical; we therefore need to do the maximum of the work off line.

Our approach consists in three steps: modeling the context (using three resources), preparing and storing off line computations, and using them to provide a fast ranking of semantic associations at request time.

3.1. Context representation

The **upper domain ontology** (UDO) describes the organization activity domains by relating the most important concepts to one another. We name these concepts *semantic descriptors* (SD). They may be added an XML[4] attribute *related resource* to refer an ontology part[5]. This means that the ontology part is to be interpreted as depending on the (sub-)domain described by the SD. One can also add an attributes *relevant task* linked up to a concept standing for a specialized task, to signify that actors of the (sub-)domain may perform such tasks. Those tasks are referred to in the Task hierarchy by the generic tasks they specialize.

The **task hierarchy** describes tasks in terms of *used tools*, *exploited material*, *needed competence* using XML attributes of that even names, and linked to semantic descriptors. *Specific tasks* may be related to ontology parts, by a *relevant resource* attribute, meaning that the ontology part is interpreted as a relevant resource for performing the task.

The **user profile** is a lightweight hierarchy with three main branches: "general data" where are stored every informations necessary to identify the users when they connect, and allow users to recognize their colleagues; "activity domains" where the users refer the SD corresponding to their activity domains; "tasks" where they refer the TH specific tasks they are used to perform.

3.2. Off-line preparation algorithm

The system computes a semantic similarity measure between OWL ontologies concepts and SD, as well as user tasks. This information is then stored in a database, so as to be able to access to it quickly at request time.

[1]Web Ontology Language, http://www.w3.org/2004/OWL/
[2]World Wide Web Consortium, http://www.w3.org/
[3]see Jena Semantic Web Framework, http://jena.sourceforge.net/
[4]Extensible Markup Language (XML), http://www.w3.org/XML/
[5]An ontology part is constituted by a concept and all its sub-concepts.

For each ontology concept c, we store the $domains(c)$ list of SD that refer any part it is included in. We then associate to each SD a c-dependent likelihood (Eq. 1). Semantic descriptors that do not exceed a given threshold are given the likelihood 0.

We use the notation: $\mathcal{O}$ for the UDO, sd for the semantic descriptors, c for ontology concepts, and t for tasks. The similarity measure *sim* used is asymmetrical, distance-based, and so that $sim(sd_i, sd_j) < sim(sd_j, sd_i)$ if sd_j subsumes sd_i (if c is related to sd_i, there is no evidence that it is also related to a more general concept sd_j). The first equation term is a correction value depending on the concept level of abstraction. It is based on the intuition that the more abstract the domain the user is interested in, the more probable it refer concepts that the user is not really interested in.

$$likelihood_{c,\mathcal{O}}(sd) = \frac{depth_{\mathcal{O}}(sd)}{\max_{i|sd_i \in \mathcal{O}} depth_{\mathcal{O}}(sd_i)} \times \max_{i|sd_i \in domains(c)} sim(sd, sd_i) \quad (1)$$

In the same way, for each ontology concept c we compute a list $tasks(c)$ of tasks that refer any part it is included in. We associate to each task the likelihood (Eq. 2):

$$likelihood_{c,\mathcal{O}}(t) = \sum_{i|t_i \in tasks(c)} sim(t_i, t) \quad (2)$$

3.3. *Behavior at request time*

Requests sent to the ranking system are all composed of a concept from a OWL ontologies, the user's identifier, and optionally the task that she/he wants to perform. The ranking system retrieves corresponding mappings in the AO, and information from the user profile. For each semantically related concept,it retrieves *likelihood* values for relevant domains and tasks from a database. We herein describe how the ranking system computes these values to obtain an unique value for each concept. This value will serve to rank the concepts in a contextualized manner.

Let a request be made on any user's behalf. Let us name the user's profile $\mathcal{P}$, and the request concept c_0. For each concept c we compute the $sim(c_0, c)$ value. We remove from the ranking the ontology concepts for which SD presents in the user profile are associated to a likelihood of 0. We compute for each ontology concept a measure of its adaptation (Eq. 3) to the profile and to the original concept.

$$adaptation_{c_0,\mathcal{O}}(c, \mathcal{P}) = \max_{i|sd_i \in \mathcal{P}} (likelihood_{c,\mathcal{O}}(sd_i) \times likelihood_{c_0,\mathcal{O}}(sd_i)) \quad (3)$$

The *usefulness* of the concept depends on the presence of related tasks in the user profile. It depends as well on whether there are tasks that are related to both the concept and the original concept.

$$usefulness_{c_0,\mathcal{O}}(c, \mathcal{P}) = \max_{i|t_i \in \mathcal{P}} (likelihood_{c_0}(t_i) \times likelihood_c(t_i) \times sim(c_0, t_i))$$

Finally, we rank the concepts depending on the value:

$$adaptation_{c_0,\mathcal{O}}(c, \mathcal{P}) \times usefulness_{c_0,\mathcal{O}}(c, \mathcal{P}) \times sim(c_0, c)$$

4. Functioning of the ranking system

This section shows an example of how the ranking module works, using three ontologies from the construction domain. The first one classifies a list of enterprises (Fig 1), the second one describes the domain of an organization specialized in reinforcing concrete (Fig. 2), and the third one is an ontology of concrete Fig. 3. An articulation ontology is built from these ontologies (Fig. 4).

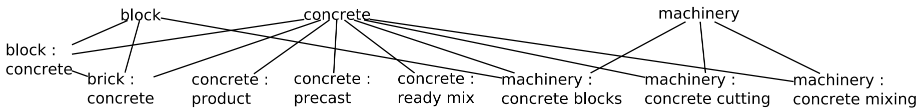

Figure 1. ontology yellowpages

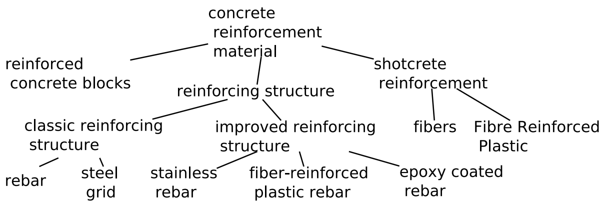

Figure 2. ontology reinforcing concrete

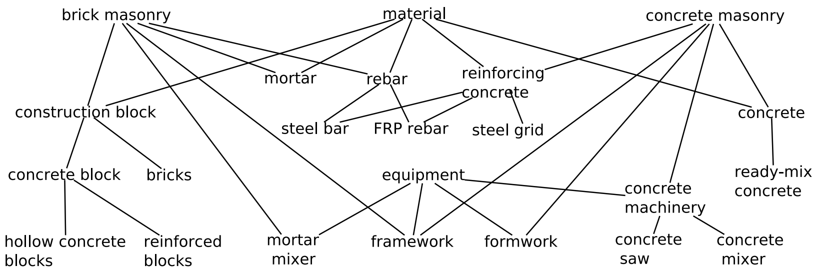

Figure 3. ontology concrete encyclopedia

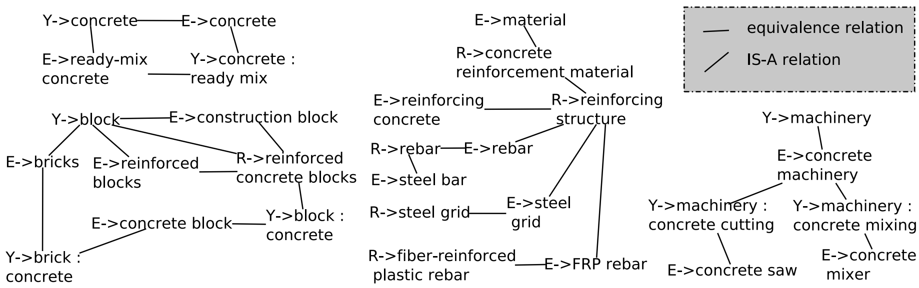

Figure 4. part of the articulation ontology

4.1. The contextual resources

We based our upper domain ontology (Fig. 5) on the UDC outline [6]. Our methodology to construct it was to first isolate the different principles of construction (here *brick masonry* and *concrete masonry*) that utilize different methods and tools. Then, we separated the concepts depending how they are employed (*material* and *equipment*). Finally, the most general concepts are chosen to be SD, and represented in the figure.

To construct the task hierarchy, we first summed up the masonry main task in a single verb (*build*), and we developed in by asking the question "how". We did not represent all the possible tasks here, for lack of place. Tasks are represented by a verb, plus a qualified direct object when necessary (e.g. to assemble by sticking, we need to have a *sticking* substance). Then, we linked these generic tasks with specialized tasks, which we put also in the UDO, so as to link them up to the concepts they are semantically related to. Finally, we inserted links to ontology concepts.

Users' *profile* can be based on a domain-specific default profile. It is a repository where the users describe their work environment and detail their work tasks. They may define public parts, so as to share knowledge with their co-workers.

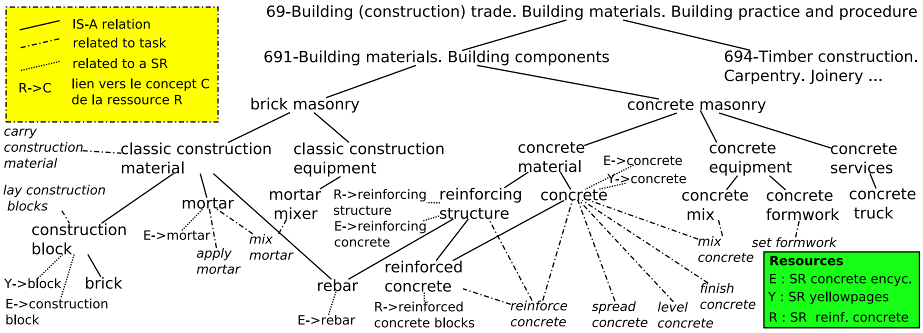

Figure 5. Upper domain ontology

Let Tom be a mason, and let him describe his domains of competencies as being *691-Building materials. Building components* and *693-Masonry and related building crafts*. His profile groups general data to identify him, and references his domains and tasks.

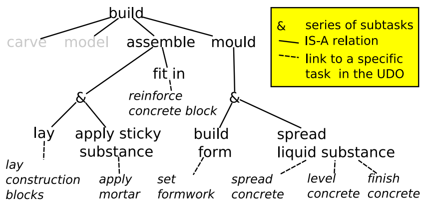

Figure 6. part of the Task hierarchy

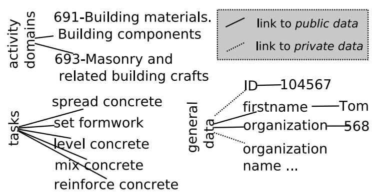

Figure 7. Tom's profile

[6]UDC consortium, see http://www.udcc.org/

4.2. Tom's request

Let Tom query the system with the concept c_0, *reinforcing concrete*, from the ontology "concrete encyclopedia".

The system retrieves all related semantic relationships, from different domains. To reduce complexity, we have not represented the related relationships defined on resources from domains such as *694-Timber construction. Carpentry. Joinery, Heating, ventilation and air conditioning of buildings*, etc.

The system filters out the concepts that correspond to the user's selected domains. We thus have: c_0 *is equivalent to* $R \rightarrow$ reinforcing structure, c_0 *generalizes* $E \rightarrow$ rebar, $R \rightarrow$ rebar, $E \rightarrow$ steel grid, $R \rightarrow$ steel grid, $E \rightarrow$ FRP rebar, $R \rightarrow$ fiber reinforced plastic rebar, $E \rightarrow$ steel bar; c_0 *is a* $R \rightarrow$ concrete reinforcment material, c_0 *is closely related to* $R \rightarrow$ reinforced concrete block, $E \rightarrow$ reinforced block.

Tom's profile refers to specific tasks as *spread concrete* and *reinforce concrete*. Only the latter is significant for the concept considered. The order of relationships is not modified, but the three last concepts are no more considered, the ranking is now: c_0 *is equivalent to* $R \rightarrow$ reinforcing structure, c_0 *generalizes* $E \rightarrow$ rebar, $R \rightarrow$ rebar, $E \rightarrow$ steel grid, $R \rightarrow$ steel grid, $E \rightarrow$ FRP rebar, $R \rightarrow$ fiber reinforced plastic rebar, $E \rightarrow$ steel bar.

Tom's request indicates that he is interested in performing the task *assemble*. Specific subtasks are *lay construction blocks* and *apply mortar*. Thus, concepts from the yellowpages ontology and with $Y \rightarrow$ block as an ancestor, or concepts from the concrete encyclopedia ontology with $R \rightarrow$ construction block as ancestor are considered as particularly relevant. Finally, the relationships c_0 *is closely related to* $R \rightarrow$ reinforced concrete block, $E \rightarrow$ reinforced block are returned as most probable, followed by c_0 *is equivalent to* $R \rightarrow$ reinforcing structure, c_0 *generalizes* $E \rightarrow$ rebar, $R \rightarrow$ rebar, $E \rightarrow$ steel grid, $R \rightarrow$ steel grid, $E \rightarrow$ FRP rebar, $R \rightarrow$ fiber reinforced plastic rebar, $E \rightarrow$ steel bar.

5. Conclusion

The approach presented in this paper consists in classifying resources depending on domains and tasks, and in using this even classification to rank concepts according to a user request: first, by filtering out the concepts defined in resources that correspond to user's domains; second, by sorting them depending on their usefulness for the user's current task.

The originality of the approach resides both in the proposal of an user-adapted semantic similarity measure to rank concepts and in the attempt to consider work tasks as a means to sort concepts depending on their usefulness.

We are now working on a prototype that implements our approach. As prospects for the future, we intend to improve our semantic similarity measure, in order to take into consideration the granularity differences between parts of the given ontologies.

Acknowledgments

We would like to thank Celson Lima and Catarina Ferreira Da Silva from the CSTB for their support and help.

References

[1] C. Ferreira Da Silva , L. Médini, S. Abdul Ghafour, P. Hoffmann, P. Ghodous: Semantic Interoperability of Heterogeneous Semantic Resources, in Proceedings of the International Workshop on Database Interoperability (InterDB 2005), Elsevier *Electronic Notes in Theoretical Computer Science*, **150** (2), 2006, pp 71–85.

[2] J. McCarthy: Notes on Formalizing Context, in Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence, Chambéry, France, 1993. URL: http://www-formal.stanford.edu/jmc/context3/context3.html

[3] B. Kokinov:A Dynamic Approach to Context Modelling, in P. Brezillon and S. Abu-Hakima (Eds), Working Notes of the IJCAI-95 Workshop on “Modelling Context in Knowledge Representation and Reasoning”, Montreal (Canada), 1995.

[4] P. Brezillon:Context in Artificial Intelligence: I. A Survey of the Literature, *Computers and artificial intelligence*, **18**(4), 1999, pp 321–340.

[5] A.K. Dey:Understanding and Using Context, in Personal Ubiquitous Computing ,**5** (1) ,London, UK, 2001, pp 4–7.

[6] R. Rada, H. Mili, E. Bicknell, M. Blettner:Development and application of a metric on semantic nets, *IEEE Transaction on Systems*, Man and Cybernetics, **19** (1), 1989, pp 17–30.

[7] Z. Wu, M. Palmer:Verb semantics and lexical selection, in Proceedings of the 32nd annual meeting of the Association for Computational Linguistics, 1994, pp 133–138.

[8] P. Resnik: Using information content to evaluate similarity in a taxonomy, in Proceedings of the 14th joint conference in Artificial Intelligence, 1995.

[9] M. A. Rodriguez, M. J. Egenhofer: Putting similarity assessments into context: matching functions with the userŠs intended operations, in Proceedings of the 2nd international and interdisciplinary conference on modeling and using context CONTEXTŠ99, LNAI 1688, 1999, pp 310–323.

[10] J. F. Roddick, K. Hornsby and D. de Vries:A Unifying Semantic Distance Model for Determining the Similarity of Attribute Values., ACSC, 2003, pp 111–118.

[11] B. Aleman-Meza, C. Halaschek, I. B. Arpinar, and A. Sheth: Context-Aware Semantic Association Ranking, in Int. Conf. on Semantic Web and Databases, Germany, 2003, pp 33–50.

[12] W. Nejdl, R. Paiu: Desktop Search - How Contextual Information Influences Search Results and Rankings, in Proceedings of the ACM SIGIR 2005 Workshop on Information Retrieval in Context (IRiX), August 2005, Salvador, Brazil.