Christophe Debruyne. On the Social Dynamics of Ontological Commitments. In Robert Meersman, Tharam S. Dillon, and Pilar Herrero, editors, OTM Workshops, volume 6428 of LNCS, pages 682-686. Springer, 2010

On the Social Dynamics of Ontological Commitments

Christophe Debruyne

Semantics Technology and Application Research Lab (STARLab)
Department of Computer Science
Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium
chrdebru@vub.ac.be

Abstract. The aim of my thesis is to provide a solution in which any stakeholder of a particular community (with a specific goal) can contribute to the ontology construction process, making the contribution of the "unproductive" long tail of the community relevant. Members in a community will describe their view and maintain a dialogue in natural language. I believe that granting the community first-class-citizenship within ontological commitments by (i) mapping those natural language descriptions and dialogues to formal descriptions and decisions within the ontology engineering process and (ii) exploiting the existing application commitments to that ontology improves the quality of ontological commitments by truly representing the community and their latest requirements.

1 Introduction

Modeling ontologies for the Semantic Web (SW) [4] is far from trivial: providing more rules that are important for effective and meaningful interoperation between applications may (and will) limit the generativity of an ontology [18] and lightweight ontologies that hold none or few domain rules, are not very effective for communication *between autonomously developed software agents*. Thanks to the Linked Data[1] (LD) initiative, we have access to billions of RDF triples ready to be exploited and even though the vast amount of data publicly available indicated a success in community effort, the absence of the so-called "killer application" indicates an issue. In LD, the primary reference scheme for concepts and their instances is a URI instead of a (more conceptually correct) reference scheme based on the attributes (literal or non-literal) of a concept. A conceptual reference scheme is necessary for applications with a specific purpose and is the reason why current popular SW technologies have a difficult uptake in "real" business.

My research aims to solve this issue by granting the community a more prominent role in the construction of ontological commitments on what and how things should be communicated. As a natural consequence, natural language (NL) will be used to reach consensus within community members for which we can draw inspiration from proved database techniques with groundings in linguistics [11,21]. As online

[1] `http://www.linkeddata.org/`

communities' member-contribution follows the 80-20 Rule[2], stating that 80% of the people are responsible for only 20% of the contribution. The technology used in those networks aims to exploit the 80% "unproductive" long tail and make their contribution significant.

2 Research Questions

What are the characteristics of a tri-sortal Internet? The Web is currently shifting from an information retrieval medium to a participatory medium spreading widely involving three types of actors: humans, computer systems and businesses. Three parallel and interconnected evolutions are simultaneously taking place on that Web, though unsynchronized [15]: (i) *Technology*, as exemplified by the SW; (ii) *Social forces* as manifested in the Social Web and (iii) *Economical forces*, the Internet now being the medium of choice for most content-based interaction of businesses or Enterprise 2.0. This shift changes the way how we should look at the Web.

Why is there a need for empowering organized communities in creating ontological commitments on a tri-sortal Internet? The introduction of business as an additional dimension motivates the need of making concepts such as group, community, etc. explicit. Community involvement is essential for interoperability as well as facilitating the uptake of LD. The LD initiative however relies on the URI mechanisms of RDF(S) to represent data and these are difficult to understand by non-technical people. Enabling communities to develop and maintain a representation of their (business) world needs a methodology since reaching a common agreement between many stakeholders proves to be difficult [7]. Appropriate methodologies for this can learn from database modeling principles.

What is a viable approach for empowering organized communities on a tri-sortal Internet? Social Web applications such as wikis and Content Management Systems (CMS) have already proven to be an appropriate tool for community participation on the Social Web and on-line communities will become the environment for meaning agreement, as it is key to achieve interoperability between systems and businesses. The result of such process of determining meaning by community agreement is called *Social Semantics*. This process hints a certain duality as two distinct and coexisting perspectives are needed. The first is the *Human perspective* characterized by high level reasoning about the shared concepts by and between humans in NL (through, for instance, dialogue). The second is the *System perspective* where agreement is made on the vocabulary used, the data access and room for simple, low level reasoning. This dual perspective forms the essence of *hybrid ontologies* [15] where concepts on the one hand are circumscribed linguistically and (mostly) declaratively by agreement within (human) communities, and on the other hand identified formally (and unambiguously) for use in computer-based information systems. Answering this question means looking for appropriate methods and tools for hybrid ontology engineering (OE).

How can (existing) ontological commitments guide the dialogue between the community members? Ontological commitments can be seen as software objects

2 Which states that for many events, roughly 80% of the effects come from 20% of the causes.

either manifested as agents or services that use these mappings to add semantics to their data. Even though a methodology supports an ontology to co-evolve with the communities' interoperability-requirements, doing so in an automated way still remains an open question. Members of a community might enter an observation while working on an ontology that might be true for their application, but not for the applications of other stakeholders. Counterexamples for such an observation result in the refusal of that observation, refinement of the ontology or the detection of mistakes in the data sets. I therefore investigate a method and tool to support an OE process by testing hypotheses on annotated data sets.

3 Related Work

Ontology Engineering. Before communities can use information and interoperability between information systems is established, a consensus on an ontology needs to be achieved among its different stakeholders. Various methodologies for OE have been developed to reach that consensus such as DOGMA-MESS [7], DILIGENT [17] and HCOME [12]. Application symbols are mapped onto concepts in that ontology once the community reaches an agreement.

DOGMA and DOGMA-MESS. DOGMA [14] is an ontology approach having some characteristics that make it different from traditional ontology approaches such as its groundings in the linguistic representations of knowledge and the methodological separation of the domain- and application-conceptualization [18]. The knowledge building blocks – called *lexons* [14] – only need in principle to express "plausible" facts (as perceived by the community of stakeholders) in order to be entered into the *Lexon Base*, a repository containing large sets of such lexons. A lexon is formally described as a 5-tuple ⟨*γ, headterm, role, co-role, tailterm*⟩, where γ is an abstract *context identifier* pointing to a resource such as a document on the Web. The context identifier is assumed to identify unambiguously (to human users at least) the concepts denoted by the term and role labels. The *Commitment Layer* contains ontological commitments that use a selection of lexons to annotate applications and specify constraints defining the use of the concepts in the ontology. DOGMA distinguishes two types of ontological commitments: *community commitments* and *application commitments*. The first denotes a meaningful selection of lexons and constraints that capture well the intended semantics of a domain. The latter extends the community commitments with mappings describing how one individual application commits to the ontology using Ω-RIDL [20].

Formalizing Dialogue. Knowledge management involves communication among loosely structured networks and communities of people with (complex) social practices and relationships that are happening in a particular context. Dialogue is one example for creating and communicating knowledge that can be supported by tools. Most work on the formalization of dialogue in computer science, however, is built around the multi-agent community [9,5].

Testing Hypotheses against Data Sets. By our knowledge, no-one has so far done work on incrementally extending ontologies by testing hypotheses to annotated databases, most existing work relates to manually or (semi-)automatically [19,1] mapping relational databases to RDF(S) and OWL. The latter often looks at the automatic transformation of database content and the schema. In [16], OWL is extended to cope with aspects of relational databases while reasoning over the data. In ontology matching [10], conflicting information resulting from schema/ontology mappings is used to improve the mappings. The annotations will be based on Ω-RIDL, whose application mappings (see Table 1) can be used to create the necessary queries to test the database for counterexamples.

1	`Person with / of First Name.`
2	`map "tblPerson"."fname" on First Name of Person.`
3	`Person has at most 1 First Name.`

Table 1. Examples of controlled sentences in Ω-RIDL: (1) depicts a lexon, (2) shows the mapping of a field in a table on a term of a lexon and (3) shows a constraint on that lexon.

Wikis for Ontology Engineering. Wiki technology has been put forward as a mean to reach agreement and share knowledge about different subjects over the past decade. The advantage of Wiki technology is that anyone can add content without much technical knowledge. Wiki technology has been adapted in the field of OE to enable non-technical users to create, visualize and maintain ontologies [3] or to semantically annotate the content [2,13].

4 Material and Setting

STARLab offers a dynamic and collaborative environment with quite a few international contacts and participates in a number of European and local projects in which my research is and will be applied, e.g., TAS[3] (Trusted Architecture for Securely Shared Services, EU FP7 216287). I have furthermore the unique opportunity to conduct experiments during the practical session of the Open Information Systems course taught to (primarily) 1st year MSc in Computer Science. The data obtained during such experiments already gave some interesting results published in [6]. The experiments with the students will be used to draw some first conclusions (e.g., via surveys) and its results as a benchmark for deployment within projects.

5 Conclusions and Future Work

The billions of triples given to us by LD are of little use for business applications as URIs are the primary reference scheme and some constraints are not possible nor imposed (e.g., identification constraints on multiple attributes). Popular Semantic Web technologies create a barrier for non-technical stakeholders, leaving them behind. My research aims to empower communities on the Tri-sortal internet by expressing their knowledge and thoughts in NL.

To this end, two prototypes are currently under development: GOSPL [8] and Ω-DIPPER. GOSPL stands for Grounding Ontology with Social Processes and Natural Language and tries to bridge the gap between the formal and informal descriptions of concepts. It is built on top of DOGMA and is currently being used within TAS[3] to allow end users (e.g., security and privacy experts) to easily develop conceptual models to provide security policy interoperability. Ω-DIPPER is a tool for testing hypotheses against annotated relational databases. Counterexamples found by Ω-DIPPER trigger various OE processes and guide the dialogue between the community members. The two next steps in my research are the formalization of dialogue and the social dynamics of commitments.

References

1 Astrova, Kalja: Automatic transformation of SQL relational databases to OWL ontologies. In: WEBIST (2), INSTICC Press (2008) pp. 131–136

2 Auer, Bizer, Kobilarov, Lehmann, Cyganiak, Ives: Dbpedia: A nucleus for a web of open data. In: ISWC+ASWC 2007. Springer (2008) pp. 722–735

3 Auer, Dietzold, Riechert: Ontowiki - a tool for social, semantic collaboration. In: ISWC 2006, Springer (2006) pp. 736–749

4 Berners-Lee, Hendler, Lassila: The Semantic Web. Scientific American **284**(5) (2001) pp. 35–43

5 Besnard, Doutre, Hunter (eds.): Computational Models of Argument: Proceedings of COMMA 2008, IOS Press (2008)

6 De Leenheer, Debruyne, Peeters: Towards social performance indicators for community-based ontology evolution. In: CK 2008, collocated with ISWC 2009, CEUR-WS (2009)

7 de Moor, De Leenheer, Meersman: DOGMA-MESS: A meaning evolution support system for interorganizational ontology engineering. In: ICCS 2006, Springer (2006) pp. 189–203

8 Debruyne, Reul, Meersman: GOSPL: Grounding ontologies with social processes and natural language. In: ITNG 2008, IEEE Computer Society (2010) pp. 1255–1256

9 Dunne, Bench-Capon (eds.): Computational Models of Argument: Proceedings of COMMA 2006, IOS Press (2006)

10 Euzenat and Shvaiko: Ontology matching. Springer (2007)

11 Halpin: Information Modeling and Relational Databases. Morgan Kaufmann, San Francisco, CA, USA (2008)

12 Kotis, Vouros: Human-centered ontology engineering: The HCOME methodology. Knowl. Inf. Syst. **10**(1) (2006) pp. 109–131

13 Krötzsch, Vrandecic, Völkel, Haller, Studer: Semantic wikipedia. J. Web Sem. **5**(4) (2007) pp. 251–261

14 Meersman: Semantics ontology tools in information system design. In: ISMIS 1999, Springer (1999) pp. 30-45

15 Meersman, Debruyne: Hybrid ontologies and social semantics. In: DEST 2010, IEEE Press (2010)

16 Motik, Horrocks, Sattler: Bridging the gap between owl and relational databases. Web Semantics: Science, Services and Agents on the WWW **7**(2) (2009) pp. 74–89

17 Pinto, Staab, Tempich: Diligent: Towards a fine-grained methodology for distributed, loosely-controlled and evolving engineering of ontologies. In: ECAI 2004, IOS Press (2004) pp. 393-397

18 Spyns, Meersman, Jarrar: Data modelling versus ontology engineering. SIGMOD Record Special Issue **31 (4)** (2002) pp. 12–17

19 Tirmizi, Sequeda, Miranker: Translating SQL applications to the semantic web. In: DEXA 2008, Springer (2008) pp. 450–464

20 Trog, Tang, Meersman: Towards ontological commitments with Ω-ridl markup language. In: RuleML 2007, Springer (2007) pp. 92-106

21 Wintraecken: The NIAM Information Analysis Method, Theory and Practice. Kluwer Academic Publishers (1990)