Achieving Maturity: The State of Practice in Ontology Engineering in 2009

Elena Simperl[1], Malgorzata Mochol[2], Tobias Bürger[1], and Igor O. Popov[1]

[1] Semantic Technology Institute (STI) Innsbruck, University of Innsbruck, Austria
[2] Netzbasierte Informationssysteme (NBI), Freie Universität Berlin, Germany
{elena.simperl,tobias.buerger,igor.popov}@sti2.at,
mochol@inf.fu-berlin.de

Abstract. In this paper we give an account of the current state of practice in ontology engineering (OE) based on the findings of a 6 months empirical survey that analyzed 148 OE projects. The survey focused on process-related issues and looked into the impact of research achievements on real-world OE projects, the complexity of particular ontology development tasks, the level of tool support, and the usage scenarios for ontologies. The main contributions of this survey are twofold: 1) the size of the data set is larger than every other similar endeavor; 2) the findings of the survey confirm that OE is an established engineering discipline w.r.t the maturity and level of acceptance of its main components, methodologies, etc. whereas further research should target economic aspects of OE and the customization of existing technology to the specifics of vertical domains.

1 Introduction

Semantic technologies are entering mainstream IT and ontologies, as a core building block of the semantic technology stack, are used to tackle a number of important aspects of modern IT, from enabling interoperability to sharing knowledge. The achievements of the ontology engineering (OE) community in the last over 20 years form a solid basis for the usage of ontologies in all technical contexts across various domains. OE has already found many applications e.g. in eCommerce while for other domains initiatives such as VoCamps[1] provide an organizational framework to meet, reach a common understanding, and develop ontologies of general interest. Engineering methodologies provide process-oriented guidelines for the development and maintenance of ontologies. Numerous methods and techniques are available for extracting ontologies from other structures or resources such as text corpora[5], classifications and taxonomies[13], folksonomies[36], or database schemes[1]; and for matching, merging, and aligning ontologies[7]. Finally, OE environments such as Protégé[2] and TopBraid-Composer[3] provide a rich list of features supporting particular tasks within the ontology lifecycle.

[1] http://www.vocamp.org
[2] http://protege.stanford.edu
[3] http://www.topquadrant.com/topbraid/composer/

R. Meersman, T. Dillon, P. Herrero (Eds.): OTM 2009, Part II, LNCS 5871, pp. 983–991, 2009.

In this paper we present an update on the state of the art in OE in 2009 based on an empirical survey performed between Oct. 2008 and March 2009 that collected data from 148 OE projects from industry and academia in order to give an account of the current OE practice and the effort involved in these activities. Just as our work from 2006[25], the survey focused on process-related rather than modeling issues; it analyzed the impact of research achievements on real-world OE projects, the complexity of particular ontology development tasks, the tool support, and the application scenarios of ontologies. The main contributions of this survey compared to other related work are twofold: 1) the size of the data set is by far larger than other similar endeavor,2) the findings of the survey confirm that OE is by now an established engineering discipline, providing the full range of methodologies, methods, techniques, and tools that allow for real-world projects to be feasibly undertaken, to some extent without external OE consultancy.

The paper is organized as follows: After a brief overview of the OE field (Sec. 2) we present the design of our survey and discuss its most important results (Sec. 3). Sec. 4 provides a summary of existing studies published in the OE literature, and explains their relationship to our research. Sec. 5 summarizes the main findings of the survey.

2 Ontology Engineering in a Nutshell

Ontology engineering (OE) is formally defined as “the set of activities that concern the ontology development process, the ontology life cycle, and the methodologies, tools and languages for building ontologies”[10]. OE methodologies[10,31] can be classified into two main categories, depending on the setting in which they are applied: *centralized OE* [2,8,30] suitable for the development of ontologies for a specific purpose within an organization; the engineering team is concentrated in one location and communication in team occurs in F2F meetings, and *decentralized OE*[16,27] more relevant in the Semantic Web context or other large-scale distributed environments; the team members are dispersed over several locations and affiliated to different organizations, communication is typically asynchronous and the ontology provides a lingua-franca between different stakeholders or ensures interoperability between machines, humans, or both.

Methodologies decompose the OE process in varied steps with different activities. [10] distinguishes among: *management (pre-development)* covering the organizational

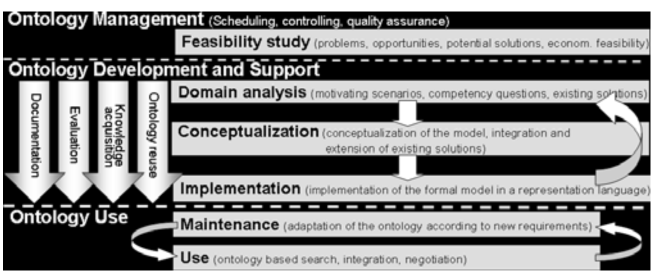

Fig. 1. Main Activities in Ontology Engineering

setting of the overall process, *development* containing domain analysis, conceptualization, and implementation, *support* including knowledge acquisition, evaluation, reuse and documentation that are performed in parallel to the core development, and *usage (post-development)* covering maintenance and use (cf. Fig. 1). Methodologies also define the roles of the individuals involved in the OE process: *domain experts* providing knowledge w.r.t the given domain, *ontology engineers* with expertise e.g. in knowledge representation or ontology tools, and *users* applying the ontology for a particular purpose.

Recent OE research investigates how to apply Web 2.0 to the development of community-driven ontologies. Approaches based on Wikis[34], tagging[3] or casual games[29] are some of the most prominent examples. Methodologies for ontology reuse[9,26] or ontology learning[19,28] guide the support activities of the ontology life cycle.

3 Our Survey

The aim of our survey was to: i) capture the basic OE understanding of semantic technology adopters; ii) give an account of the state of OE practice as of 2009; iii) assess the level of support provided by OE methodologies, methods and tools;& iv) suggest directions for further research. Through its size and range of subjects (data from 148 projects) the survey gives a comprehensive and representative overview of the current state and can be seen as an indicator of the positive trends in the OE. The data was gathered through F2F or telephone interviews (approx. 50% of the projects), and an online questionnaire. The respondents were representative for the OE community: IT practitioners, researchers and experts from various disciplines, affiliated to industry or academia, participating in the last 3 to 8 years in ontology development projects in areas such as Information Systems, eCommerce, multimedia. More than 95% of the projects were carried out in Europe, whilst nearly 35% originated from industry parties and "industrial" ontologies were mostly used in commercial IT solutions. Most of the ontologies were either domain or application, whereas few of them were core ontologies. The size of the ontologies varied from 60 to 11 million entities. The representation language of choice was OWL DL (30%), followed by WSML DL and WSML Flight (around 10% each) and RDF(S) (9%). The effort of the OE projects varied from 0.02 to 156 PMs.

The survey was supported by an online questionnaire[4] consisting of 38 open-ended and scaled questions divided into: i)*general aspects of the OE project*, including the size of the resulting ontology, its purpose and the development costs, ii) questions regarding the *ontology development phases*, such as domain analysis, conceptualization, implementation, and evaluation, iii) *characteristics of the engineering team* engaged in the project, and iv) questions about the *software used* to support and guide the process.[5]

General Issues: The survey clearly pointed out that the use of methodological support for developing ontologies varied from project to project. As with previous findings, some IT professionals and researchers did not perceive ontology development as a systematic process which should be performed according to a predefined methodology,

[4] The questionnaire is available at `http://ontocom.sti-innsbruck.at`
[5] A detailed structure of the questionnaire can be found in our previous work in [25].

nevertheless, the process was largely inline with the general recommendations found in the literature. On average, use of some methodology was observed in 1 out of 9 projects though, in the case of more challenging, complex or specialized ontology development projects, the ratio was 50%. Since process-driven methodologies are used in 1 out of 2 projects, in which assistance to the OE team is expected to be essential, we argue that this is a clear indicator for level of maturity achieved by OE and, as in other engineering disciplines, that an increasing number of projects resort to predefined methodological support confirms the fact that OE is an established discipline. This finding is different from the ones of previous surveys published two to more years ago. In terms of directions for improvement, participants suggested that project settings in which requirements/domain analysis and evaluation needs run high mandate domain-specific adaptations of generic methodologies. This might be beneficial for development of ontologies with broad coverage or involving non-common-sense knowledge.

Process, Personnel and Project Issues: The structure of the survey assumed the activities breakdown of OE processes (cf. Sec. 2) which proved to match to a satisfactory extent to the way the surveyed projects carried out the process. The interviews emphasized, however, some discrepancies between i) the complexity of particular activities as perceived by OE practitioners, ii) the significance of these activities as measured in terms of their impact on the total development costs, and iii) the level of maturity achieved by the OE community w.r.t methods and tools supporting these activities. To investigate the relationship between the OE aspects, their interdependencies and impact on the development costs, we performed a correlation analysis that provides an overview of the importance of each aspect, and assists in identifying those aspects whose impact might have been underestimated so far and would require additional attention. Aspects can be positively (value between 0.1 and 1), negatively correlated (between -0.1 and -1) or independent (between -0.1 and 0.1). Overall, the outcomes of the correlation analysis were consistent with most of the interviewees feedback.

– *Correlation between OE aspects and effort:* Our findings (cf. Tab. 1) point out that activities like domain analysis, conceptualization, implementation, evaluation, level of reusability of the ontology, and documentation requirements have a well-distributed

Table 1. Correlation between OE aspects and effort in person months

OE aspect	Description	Correlation with effort
DCPLX	Complexity of the domain analysis	0.496
CCPLX	Complexity of the ontology conceptualization	0.237
ICPLX	Complexity of the ontology implementation	0.289
REUSE	Required reusability of the ontology	0.274
DOCU	Complexity of the documentation task	0.346
OEV	Complexity of the evaluation	0.362
OCAP/DECAP	Capability of the ontologists/domain experts	-0.321
OEXP/DEEXP	Expertise of the team	-0.192
PCON	Personnel continuity	-0.134
LEXP/TEXP	Level of experience with respect to languages and tools	-0.172
SITE	Communication facilities in decentralized environments	-0.168

correlation factor associated with the effort; these activities exhibit a relevant impact on the effort, at the same time indicating that no activity plays a predominant role.

Domain analysis: Out of the 6 positively correlated factors, DCPLX had the highest impact on the total effort, achieving a significantly higher correlation value. This is an assessment of the time-consuming nature of the knowledge acquisition, which was also confirmed by our interviews and previous surveys. Many interviewees questioned the utility of tools for this activity, which were perceived as too generic especially when it came to ontologies developed for highly specialized domains such as health care, or in projects relying on end-user contributions. Our survey revealed that the majority of participants at times used a variety of knowledge elicitation techniques in an informal manner, outside the stated guidelines of a specific methodology. These techniques, complemented with detailed insights of the practices established in particular domains, could be useful to design specially targeted OE methodologies and guidelines.

Ontology evaluation: The quality of the implemented ontology remains a major concern among ontology engineers. Nevertheless, the surveyed projects seldom used any of the existing ontology evaluation methods and techniques, but relied on expert judgement. In projects in which systematic ontology evaluation practices were observed, they immediately had a significant impact on the effort. More than 50% of projects reported minor effort in formally testing the developed ontologies, 48% reported fair use of simple testing methods carried out mostly manually, only 3 projects performed extensive testing using several methods. The survey indicated a combination of manual testing and self-validation by the engineering team as the preferred and common choice in most projects. Thus, ontology evaluation plays a passive role in ontologies developed in less formal project settings, however, as evaluation practices increase with the demand for quality assurance, the associated impact on effort can be substantial.

The dominance of DCPLX and OEV indicates that any easing in these activities may result in major efficiency gains. The respondents indicated a low tool support or lack of easy to use tools for these tasks - w.r.t. the domain analysis the situation could be improved e.g. by applying automated document analysis or ontology learning approaches.

The impact of personnel-related aspects suggests that more training programs in the OE area, better collaboration support, and an improved, fine granular documentation of the decisions taken during the OE process may have positive effects. The data analysis produced counter-intuitive results for the SITE parameter: emails lowered the effort needed to build ontologies while frequent F2F meeting increased the effort significantly (F2F meetings produced more different views on the ontology and resulted in more discussions which raises the costs of ontology development).

– *Correlation between OE aspects:*[6] Personnel-related aspects (cf. Tab. 2) were shown to be positively correlated. This was obvious for questions referring to the capability and experience of the OE team: the capability was largely based on the team experience.

The software support in projects carried out by the same OE team tended to remain unchanged. When new tools were introduced, the learning period for experienced developers was much higher than for novel. High correlation was also measured between

[6] Since it is not possible to account for all possible relationships between certain OE aspects in the scope of this paper, we restrict ourselves to the more important findings.

Table 2. Correlation between Personnel-Related Aspects

	OCAP/DECAP	OEXP/DEXP
OCAP/DECAP	1	0.552
OEXP/DEXP	0.552	1
LEXP/TEXP	0.489	0.584

activities within the OE process: The correlation between evaluation and documentation was concentrated on large-scale OE projects since such projects run more extensive evaluation tests, which might lead to additional documentation effort. Domain analysis was most highly correlated with the conceptualization and implementation. In over 40% of the projects the ontology development was performed mainly by domain experts, who agreed that current editors are relatively easy to learn and utilize. This finding is different from the result of previous studies, and confirms one more time the fact that OE has reached an industry-strength level of maturity.

4 Related Work

This section gives a brief overview of the OE surveys and case studies previously published: **Analytical surveys** analyze OE methodologies from the theoretical perspective and identify open issues in this field. In the nineties the surveys focused on laying out the foundations of OE[18,23], highlighted the need for guidelines and best practices on ontology development and reuse, and fully fledged, tool-supported methodologies[11,14]. Some of these issues were still to be solved years later: a comprehensive survey published in 2003[6] identified a lack of software support for many methodologies and their limited scope as compared to the ontology life cycle. More recent surveys emphasized the integration of OE into the enterprise (business process) modeling landscape and the study of OE economics as essential issues for the adoption of OE beyond academia[31]. **Empirical surveys** focus on insights and findings derived from real-world case studies on ontology development, management or reuse. Most of such surveys reported on the application of self-developed methodologies, highlighting their advantages as compared to alternative engineering approaches[33,35]. Some surveys described the deployment and evaluation of a particular OE methodology or tool in a given context[9,17] while other reported on practical experiences in an ontology-related project[21,24]. **Case studies** evaluated several methodologies and methods w.r.t their relevance and usability, prior to their application in a particular setting, or operated the engineering process without nominally committing to existing techniques[15,32]. Additionally,[20,22] provided some practical guidelines and recommendations for developing ontology-based applications in specific sectors. **Other studies** highlighted the limited awareness of OE methodologies in commercial settings which hampered industrial adoption[4], recommended intensified promotion measures for OE methodologies and their benefits, to raise the awareness of semantic technology researchers and practitioners in this respect[25], and pointed out the need for advanced technology to cope with ontology development and maintenance in rapidly changing domains[12].

5 Conclusions

Even if industry is starting to acknowledge the technical value of ontologies, the information known about the process underlying the *development of ontologies in practice* is still very limited: The literature reports mainly on case studies which involved methodologists, while ontologies are envisioned to be developed by domain experts possessing limited to no skills in OE and recent surveys are either of analytical nature or not supported by a critical mass of adopters. The aim of this paper was to fill this gap through a study with 148 projects that developed ontologies for commercial and academic applications in a wide range of domains. This is by far the largest survey of this kind investigating the systematics, development effort and problems encountered in a significant share of the most relevant and popular OE projects ever run in the rapidly evolving semantic landscape. The main findings of the survey are: i) OE methodologies are used in projects developing large ontologies or being under critical requirements; ii) an increasing number of projects involve end-users in the development (editors seem to be well-appropriate to be used by less technology-prone users, though teams are reluctant to changing a tool environment due to the high learning curve); iii) ontology engineers need cost benefit analysis methods to determine the transition point between OE activities; and iv) with the uptake of ontology-based technology the need arises for methodologies and techniques customized for the characteristics of particular vertical domains (mainly for activities having a high impact on the development costs). All these findings confirm the fact that OE can be considered an established engineering discipline - methodologies are used in projects whose success critically depends on a systematic operation of the engineering process, whilst end-users become more involved in the development of ontologies with the help of mature ontology management tools.

Acknowledgements. The research leading to this paper was partially supported by the European Commission under the contracts FP6-027122 “SALERO’ and FP7-215040 “ACTIVE”.

References

1. Astrova, I.: Reverse engineering of relational databases to ontologies. In: Bussler, C.J., Davies, J., Fensel, D., Studer, R. (eds.) ESWS 2004. LNCS, vol. 3053, pp. 327–341. Springer, Heidelberg (2004)
2. Benjamin, P.C., et al.: Ontology capture method (IDEF5). Technical report, Knowledge Based Systems, Inc. (1994)
3. Braun, S., et al.: Ontology Maturing: a Collaborative Web 2.0 Approach to Ontology Engineering. In: Proc. of the Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007) at the 16th International WWW 2007 (2007)
4. Cardoso, J.: The Semantic Web Vision: Where Are We?. IEEE Intelligent Systems 22(5), 84–88 (2007)
5. Cimiano, P.: Ontology Learning and Population from Text: Algorithms, Evaluation and Applications. Springer, Heidelberg (2006)
6. Corcho, O., Fernández-Lopéz, M., Gómez-Pérez, A.: Methodologies, tools and languages for building ontologies: where is their meeting point? Data & Knowledge Engineering 46(1), 41–64 (2003)

7. Euzenat, J., Shvaiko, P.: Ontology Matching. Springer, Heidelberg (2007)
8. Fernandez, M., Gomez-Perez, A., Juristo, N.: Methontology: From ontological art towards ontological engineering. In: Proc. of the AAAI 1997 Spring Symposium on Ont. Engin. (1997)
9. Gangemi, A., Pisanelli, D.M., Steve, G.: Ontology integration: Experiences with medical terminologies. In: Formal Ontology in Information Systems, pp. 163–178. Press (1998)
10. Gómez-Pérez, A., Fernández-López, M., Corcho, O.: Ontological Engineering. Springer, Heidelberg (2003)
11. Grüninger, M., Fox, M.: Methodology for the design and evaluation of ontologies. In: Proc. of the IJCAI 1995, Workshop on Basic Ontological Issues in Knowledge Sharing (1996)
12. Hepp, M.: Possible ontologies: How reality constrains the development of relevant ontologies. IEEE Internet Computing 11(1), 90–96 (2007)
13. Hepp, M., de Bruijn, J.: Gentax: A generic methodology for deriving owl and rdf-s ontologies from hierarchical classifications, thesauri, and inconsistent taxonomies. In: Franconi, E., Kifer, M., May, W. (eds.) ESWC 2007. LNCS, vol. 4519, pp. 129–144. Springer, Heidelberg (2007)
14. Jones, D., Bench-Capon, T., Visser, P.: Methodologies for ontology development. In: IT & KNOWS Conference of the 15th IFIP World Computer Congress, pp. 62–75 (1998)
15. Koenderink, N.J.J.P., Top, J.L., van Vliet, L.J.: Expert-based ontology construction: A case-study in horticulture. In: Proc. of the 16th Int. Workshop on DEXA 2005 (2005)
16. Kotis, K., Vouros, G.A.: Human-centered ontology engineering: The HCOME methodology. Knowledge and Information Systems 10(1), 109–131 (2005)
17. Lau, T., Sure, Y.: Introducing ontology-based skills management at a large insurance company. In: Proc. of the Modellierung 2002, pp. 123–134 (2002)
18. Lenat, D.B.: Cyc: A large-scale investment in knowledge infrastructure. Communications of the ACM 38(11), 33–38 (1995)
19. Maedche, A.: Ontology Learning for the Semantic Web. Kluwer Academics, Dordrecht (2002)
20. Mochol, M., Simperl, E.P.B.: Practical Guidelines for Building Semantic eRecruitment Applications. In: Proc. of International Conference on Knowledge Management (iKnow 2006), Special Track: Advanced Semantic Technologies (AST 2006) (September 2006)
21. Niemann, M., Mochol, M., Tolksdorf, R.: Improving online hotel search - what do we need semantics for? In: Proc. of Semantics 2006 (Application Paper) (2006)
22. Noy, N., McGuinness, D.L.: Ontology development 101: A guide to creating your first ontology. Technical Report KSL-01-05 and SMI-2001-0880, Stanford Knowledge Systems Laboratory and Stanford Medical Informatics (March 2001)
23. Noy, N.F., Hafner, C.D.: The state of the art in ontology design: A survey and comparative review. AI Magazine 18(3), 53–74 (1997)
24. Paslaru, E.B., Mochol, M., Tolksdorf, R.: Case studies on ontology reuse. In: Priceedings of the 5th International Conference on Knowledge Management I-Know 2005 (June 2005)
25. Paslaru-Bontas, E., Tempich, C.: Ontology Engineering: A Reality Check. In: Meersman, R., Tari, Z. (eds.) ODBASE 2006. LNCS, vol. 4275, pp. 836–854. Springer, Heidelberg (2006)
26. Pinto, H.S., Martins, J.: Reusing ontologies. In: AAAI 2000 Spring Symposium on Bringing Knowledge to Business Processes, pp. 77–84 (2000)
27. Pinto, H.S., Tempich, C., Staab, S.: Diligent: Towards a fine-grained methodology for distributed, loosely-controlled and evolving engingeering of ontologies. In: Proc. of the 16th European Conference on Artificial Intelligence (ECAI 2004), pp. 393–397. IOS Press, Amsterdam (2004)
28. Simperl, E., Tempich, C.: A Methodology for Ontology Learning. In: Bridging the Gap between Text and Knowledge - Selected Contributions to Ontology Learning and Population from Text. IOS Press, Amsterdam (2007)

29. Siorpaes, K., Hepp, M.: Games with a purpose for the semantic web. IEEE Intelligent Systems 23(3), 50–60 (2008)
30. Sure, Y., Staab, S., Studer, R.: Methodology for development and employment of ontology based knowledge management applications. SIGMOD Record 31(4), 18–23 (2002)
31. Sure, Y., Tempich, C., Vrandecic, D.: Ont. Engineering Method. In: Semantic Web Technologies: Trends and Research in Ontology-based Systems, pp. 171–187. Wiley, Chichester (2006)
32. Tautz, C., Althoff, K.D.: A case study on engineering ontologies and related processes for sharing software engineering experience. In: Proc. of the International Conference on Software Engineering and Knowledge Engineering SEKE 2000 (2000)
33. Tempich, C., Pinto, H.S., Staab, S.: Ontology engineering revisited: an iterative case study with diligent. In: Sure, Y., Domingue, J. (eds.) ESWC 2006. LNCS, vol. 4011, pp. 110–124. Springer, Heidelberg (2006)
34. Tempich, C., Simperl, E., Pinto, S., Luczak, M., Studer, R.: Argumentation-based Ontology Engineering. IEEE Intelligent Systems 22(6), 52–59 (2007)
35. Uschold, M., King, M., Moralee, S., Zorgios, Y.: The enterprise ontology. Knowledge Engineering Review 13(1), 31–89 (1998)
36. Van Damme, C., Hepp, M., Siorpaes, K.: Folksontology: An integrated approach for turning folksonomies into ontologies. In: Bridging the Gep between Semantic Web and Web 2.0 (SemNet 2007), pp. 57–70 (2007)