International Journal of Computer Science and Applications

Vol. 7 No. 1, pp. 45 - 65, 2010

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

Elena Simperl
Semantic Technology Institute (STI) Innsbruck, University of Innsbruck
Technikerstrae 21a, 6020 Innsbruck, Austria
elena.simperl@sti2.at
http://members.sti2.at/∼elenas

Malgorzata Mochol
Netzbasierte Informationssysteme (NBI), Freie Universität Berlin
Königin-Luise-Str. 24-26, 14195 Berlin, Germany
mochol@inf.fu-berlin.de
http://page.mi.fu-berlin.de/mochol

Tobias Bürger
Semantic Technology Institute (STI) Innsbruck, University of Innsbruck
Technikerstrae 21a, 6020 Innsbruck, Austria
tobias.buerger@sti2.at
http://www.tobiasbuerger.com

In this paper we give an account of the current state of practice in ontology engineering based on the findings of a six months empirical survey we performed between October 2008 and March 2009 that analysed 148 ontology engineering projects from industry and academia. The survey focused on process-related issues and looked into the impact of research achievements on real-world ontology engineering 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 compared to other works in the ontology engineering community are twofold: Firstly, the size of the data set the results are grounded on is by far larger than every other similar endeavour published in the last years. Secondly, the findings of the survey confirm the fact that ontology engineering is an established engineering discipline in respect of the maturity and level of acceptance of its main components, methodologies, methods and software tools, whereas further research should target the customization of existing technology to the specifics of vertical domains, as well as economic aspects of ontology engineering.

Keywords: ontology engineering; survey; state of the art.

1. Introduction

Semantic technologies are entering mainstream IT. Major IT vendors worldwide extend their products into support for semantics, and the results reported by adopters from vertical sectors as diverse as life sciences, telecommunications, automotive, e-commerce and e-Government convincingly demonstrate the added value of technologies such as RDF[a], RDFa[b], SPARQL[c] as well as ontologies. Confronted with these bright prospects, one question that arises is about the directions of research the Semantic Web community needs to further pursue in order to ensure a sustainable impact in sectors as those previously mentioned, and to advance the field to cope with the challenges imposed by present and future Information and Communication Technology developments. Ontologies are a core building block of the semantic technology stack. As means to formalize the kinds of things that can be talked about in a system or a context, they are increasingly being used to tackle a number of important aspects of modern IT, from enabling interoperability to managing information and sharing knowledge. The achievements of the ontology engineering community in the more than fifteen years that have passed since its foundation form a solid basis for the usage of ontologies in all these technical contexts across various application scenarios and vertical sectors - methodologies provide processes-oriented guidelines for the development and maintenance of ontologies in centralized or decentralized environments; numerous methods and techniques are available for extracting ontologies from other knowledge structures or resources such as text corpora [Cimiano (2006)], classifications and taxonomies [Hepp *et al.* (2007)], folksonomies [Van Damme *et al.* (2007)], or data schemes [Astrova (2004)], and for matching, merging, and alignment ontologies [Euzenat *et al.* (2007)]; finally, ontology engineering environments such as Protégé[d] and TopBraidComposer[e] provide a rich list of features supporting particular tasks within the ontology life cycle. To date, the dominating approach to ontology engineering is grassroots- and community-driven. Mature ontologies already exist in domains such as eHealth or eCommerce. For other domains, initiatives such as VoCamps[f] provide the organizational framework for stakeholders and enthusiasts to meet, exchange ideas, reach a common understanding, and develop ontologies of general interest. Wiki-based systems such as Semantic Media Wiki[g] are extensively used to support the ontology engineering process when it comes to knowledge elicitation or structuring. Interesting is also the approach pursued in OntoGame[h], which harnesses the wisdom of the crowds within casual games to create OWL ontologies and semantic annotations of multimedia [Siorpaes *et al.* (2008)].

In a few words, the results achieved by the ontology engineering community in the last decades are of incontestable value for the large-scale uptake of semantic

[a]`http://www.w3.org/RDF/`
[b]`http://www.w3.org/TR/xhtml-rdfa-primer/`
[c]`http://www.w3.org/TR/rdf-sparql-query/`
[d]`http://protege.stanford.edu`
[e]`http://www.topquadrant.com/topbraid/composer/`
[f]`http://vocamp.org`
[g]`http://semantic-mediawiki.org`
[h]`http://ontogame.sti2.at/`

technologies. Nevertheless, their range of application in real-world projects was so far comparatively limited, despite the growing number of ontologies online available and gradual improvements of the accompanying technology. This marginal impact was shown in several recent empirical surveys and case studies. In [Paslaru-Bontas *et al.* (2006)] the authors recommended intensified promotion measures for ontology engineering methodologies and their benefits, to raise the awareness of semantic technology researchers and practitioners in this respect. Cardoso's investigations highlighted the importance of ontology engineering methodologies in commercial settings [Cardoso (2007)]. Finally, one of the main conclusions in [Hepp (2007)] was the need for advanced technology to cope with ontology development and maintenance in rapidly changing environments. A community-driven approach to ontology engineering together with improved tool support throughout the ontology life cycle could alleviate this situation. The study also argued that the more detailed and expressive an ontology is, the less accepted and useful it is likely to be for the community, partially due to the increased effort that is required to understand and apply the ontology in a given project. Tools facilitating the reuse of ontologies could have a positive effect on this situation.

1.1. *Contributions of the Article*

In this article we present an update on the state of the art in ontology engineering in 2009. The article is based on a six months empirical survey performed between October 2008 and March 2009 that collected data from 148 ontology engineering projects from industry and academia in order to give an account of the current ontology engineering practice, and the effort involved in these activities. Just as our previous work from 2006 [Paslaru-Bontas *et al.* (2006)], the survey focused on process-related rather than modeling issues. In particular it analysed the impact of research achievements such as methodologies on real-world ontology engineering projects, the complexity of particular ontology development tasks, the level of tool support for each of these tasks, and the application scenarios of ontologies. The main contributions of this survey compared to related work in the field are twofold. Firstly, the size of the data set the results are grounded on is by far larger than every other similar endeavour published in the last years. A survey of comparable scope could be the one by Cardoso published in 2007 [Cardoso (2007)]. It contains data collected from interviews with several hundreds of researchers and practitioners; however, it targeted semantic technologies in general, and did not cover ontology engineering aspects at the same level of detail as our work. Secondly, the findings of the survey confirm the fact that ontology engineering is by now an established engineering discipline, providing the full range of methodologies, methods, techniques, and software tools that allow for real-world projects to be feasibly undertaken, to some extent even without external ontology engineering consultancy. Nevertheless, some aspects still demand further investigation. Similarly to other engineering disciplines, requirements analysis remains a challenging task. In this case, improvement could be achieved if ontology engineering technology would be customized and extended to accommodate the specifics of relevant vertical sectors, based on the practices therefore available in the respective communities of interest. Complimentary, the study of costs and benefits of ontology engineering could provide a means to steer and further structure the ontology engineering process, to decide among alternative engineering strategies, and to argue in tangible terms in favour of the adoption of ontology-based technologies. A closer integration of such economic considerations into the ontology life cycle and ontology development environments will enable a new level of quality in the management of projects employing semantic technologies, which is particularly important in commercial settings. Last, but not least, with ontologies becoming more and more popular, the availability of high-quality ontologies in key application domains, as well as of industrial-strength technology for using and reusing these ontologies will be essential for a sustainable and durable impact.

1.2. *Structure of the Article*

The rest of this article is organized as follows: Section 2 gives a brief overview of the ontology engineering field in respect of the processes, approaches and technology in place, as a baseline for the design of our survey. Section 3 presents the design of our survey and the data set collected, while Section 4 discusses the most important results and their implications. Section 5 provides a summary of existing analytical and empirical studies published in the ontology engineering literature in the last years, and explains their relationship to our research. Finally, Section 6 summarizes the main findings of the survey and concludes the article.

2. Ontology Engineering in a Nutshell

Ontology engineering 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"* [Gómez-Pérez *et al.* (2003)]. During the last years ontology engineering evolved from a purely research field to real-world applications, situation which is demonstrated by the wide range of projects with major industry involvement, and by the increasing interest of small and medium enterprises requesting consultancy on this topic. This section gives an overview of some of the most important ontology engineering activities, as a baseline for the design of our survey and its results.

Ontology engineering methodologies provide guidelines for developing, managing and maintaining ontologies; recent surveys on ontology engineering methodologies are available, for instance, in [Gómez-Pérez *et al.* (2003); Sure *et al.* (2006)]. Such methodologies decompose the ontology engineering process in a number of steps, and recommend activities and tasks to be carried out for each step. The importance of a particular activity within a concrete ontology-related project depends on the characteristics of the environment in which the ontology is to be used, the complexity of the ontology to be developed, the availability of domain-relevant information sources, and the experience of the ontology engineering team, to name but a few. Orthogonally thereof, in [Gómez-Pérez *et al.* (2003)] the authors differentiate three types of activities within an ontology engineering process - *management*, *development* and *support* activities (cf. Figure 1). The first covers the organizational setting of the overall process. In particular, at pre-development time, a *feasibility study* examines if an ontology-based application, or the use of an ontology in a given context is the right way to solve the problem at hand. The second type of activities refers to classical activities such as *domain analysis*, *conceptualization* and *implementation*, but also *maintenance* and *use*, which are performed at post-development time. Ontology support activities such as *knowledge acquisition*, *evaluation*, *reuse*, and *documentation* are performed in parallel to the development activities.

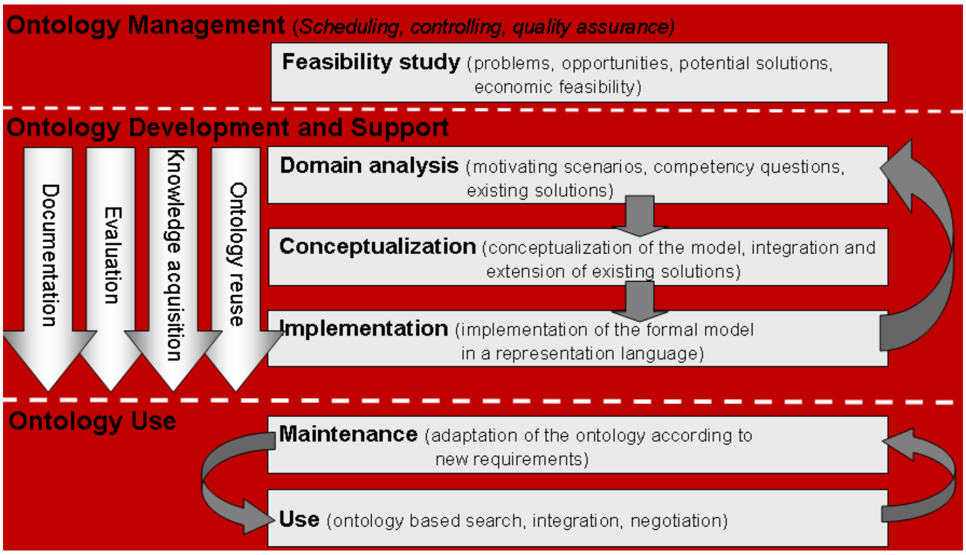


Fig. 1. Main Activities in Ontology Engineering

In addition to activities, methodologies also define the roles of the individuals and organizations involved in the project. They typically distinguish between *domain experts* providing knowledge with respect to the domain to be modeled, *ontology engineers* with expertise in fields such as knowledge representation and development tools, and *users* applying the ontology for a particular purpose. In the course of the transition of ontology engineering from research labs to real-world projects, the trend has more and more shifted towards methodological approaches which facilitate a potentially broad base of less technically skilled parties which crucially contribute to the creation and evolution of ontologies. Two main types of approaches have been proposed:

Centralized ontology engineering The ontology engineering team is concentrated in one location and communication between team members occurs in regular face-to-face meetings. This setting is primarily relevant for the "closed" development of ontologies for a specific purpose within an organization.

Decentralized ontology engineering This setting is more relevant in the Semantic Web context or in other similar large-scale open, distributed environments. The ontology engineering team is composed of stakeholders dispersed over several geographical locations and affiliated to different organizations. Communication within the team is typically asynchronous. The ontology provides a lingua-franca between different stakeholders or ensures interoperability between machines, humans, or both.

Some of the most popular methodologies for *centralized ontology engineering* are IDEF5 [Benjamin *et al.* (1994)], METHONTOLOGY [Fernandez *et al.* (1997)], and OTK [Sure *et al.* (2002)]. IDEF5 and METHODOLOGY give an overview of the most important activities of an ontology engineering process, with a focus on ontology development. As compared to the process model introduced by these methodologies, OTK's main contribution is the integration of the overall ontology engineering process into a more comprehensive framework for the realization of knowledge management applications which rely on ontologies. [Pinto *et al.* (2004); Kotis *et al.* (2005)] address the question of consensus-building, which is particularly relevant in *decentralized ontology engineering.* Following the advent of trends such as user-generated content and the Social Web, more recent research in the field looks into how to apply Web 2.0 principles and technologies in order to facilitate the development of community-driven ontologies, and to adapt earlier ontology engineering to the requirements of the new IT landscape and to the needs of modern engineering teams [Mainz *et al.* (2008)]. Classical ontology engineering is moving towards *collaborative* approaches based on wikis [Tempich *et al.* (2007)], tagging [Braun *et al.* (2007)] or casual games [Siorpaes *et al.* (2008)]. Methodologies for ontology reuse [Gangemi *et al.* (1998); Paslaru-Bontas (2007); Pinto *et al.* (2000)] or ontology learning [Maedche (2002); Simperl *et al.* (2007)] complement the overall picture, guiding the ontology support activities of the ontology life cycle.

3. Overview of the Survey

The aims of our survey were manifold:

- to capture the basic ontology engineering understanding of semantic technology adopters;
- to give an account of the state of ontology engineering practice as of 2009;
- to assess the level of support provided by available ontology engineering methodologies, methods and tools; and

- to suggest directions for further research and development in the field.

We collected data from 148 projects, a significantly higher number than in our 2006 work (34 projects) or any other similar survey published so far. The size of the data set can be seen as an indicator of the general positive trend in the field. Through its size and the range of the subjects covered, the survey gives a comprehensive overview of the current state of practice in ontology engineering. The data was gathered through face-to-face or telephone interviews (approximately 60% of the projects), the rest via an online questionnaire. The respondents are representative for the community of adopters and developers of semantic technologies. They were IT practitioners, researchers and experts from various disciplines, affiliated to industry or academia, participating in the last 3 to 8 years in ontology engineering projects in areas as diverse as Information Systems, eCommerce, multimedia, Semantic Web services, eTourism, or Digital Libraries. More than 95% of the projects surveyed were carried out in Europe, whilst nearly 35% originated from industry parties. The ontologies developed by the industry were mostly used in commercial IT solutions. Most of the ontologies were either domain or application ontologies, whereas few of them were core ontologies. The size of the ontologies in the data set varied from 60 entities to 11 million entities (see Figure 2). The knowledge representation language of choice was OWL DL[i] (30%), followed by WSML DL[j] and WSML Flight (around 10% each) and RDF(S)[k] (9%). The effort of the ontology engineering projects varied from 0.02 to 156 person months.

The survey was supported by a self-administered online questionnaire consisting of 38 open-ended and scaled questions divided into four parts.[1] The first part covers general aspects of the ontology engineering project at hand, including the size of the resulting ontology, its scope and purpose, and the development costs. The second part refers to ontology development and support activities, such as domain analysis, conceptualization, implementation, documentation and evaluation. The third part of the questionnaire is related to the engineering team which was engaged in the project. Finally, the fourth part contains questions about the software used to support and guide the process. A description of the questions is provided in Table 1, while a more detailed presentation of the structure of the questionnaire can be found in our previous work in [Paslaru-Bontas *et al.* (2006)]; the questionnaire was used, to a large extent, in the reality check survey we conducted in 2006, whose results are reported in [Paslaru-Bontas *et al.* (2006)], minor modifications being made based on feedback received in the first round.

[i] `http://www.w3.org/TR/owl-guide/`
[j] `http://www.wsmo.org/wsml/wsml-syntax`
[k] `http://www.w3.org/TR/rdf-schema/`
[1] The questionnaire is available online at `http://ontocom.sti-innsbruck.at`.

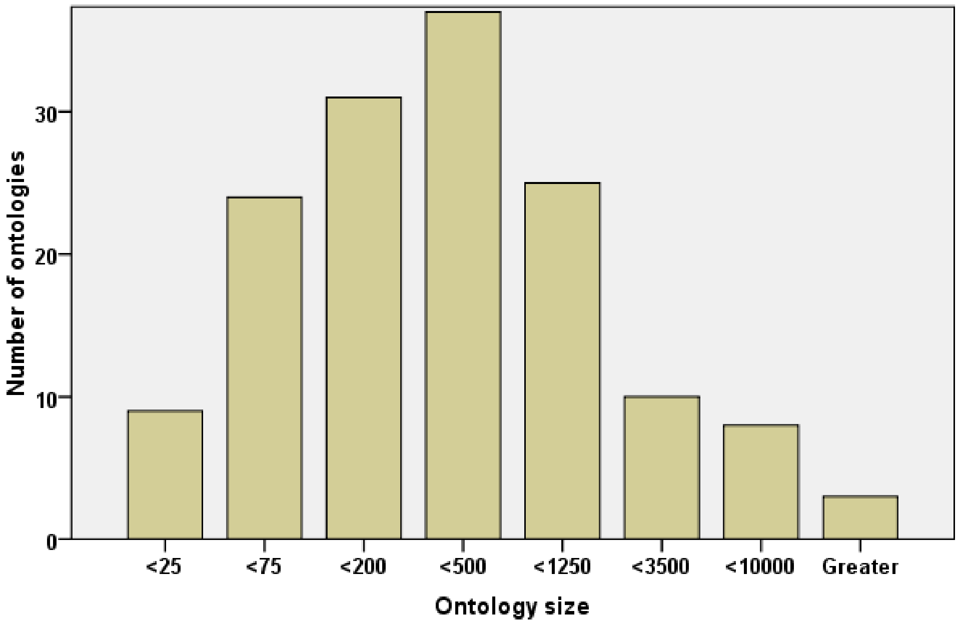


Fig. 2. Overview of the Size of the Ontologies in the Survey

4. Survey Results and Discussion

This section presents the results of the survey and discusses their implications. The section is divided into two parts. The part on general issues covers findings related to the overall ontology engineering projects. The part on process and personnel issues resorts to correlation analysis in order to identify those aspects of ontology engineering which have a high impact on the total development costs and their interdependencies.

4.1. *General Issues*

The survey pointed out that the use of methodological support for developing ontologies clearly varies from project to project. As with previous findings, some IT professionals and researchers did not perceive ontology engineering as a systematic process which should be performed according to a predefined methodology. Nevertheless, the way the overall process was carried out was largely inline with the general recommendations found in the literature in the field. Most notably, use of strict methodological guidance was available in more complex and longer ontology development projects. On average, concrete use of some methodology was observed in one out of nine projects. Lower percentages could be measured predominantly in projects developing simple ontologies or implementing requirements of less complexity. In the case of more challenging or specialized ontology develop-

No	Acronym	Description	Type
Part 1. Introduction			
1	ONTNAME	The name of the ontology	Open-ended
2	ONTONS	The namespace of the ontology	Open-ended
3	DOMAIN	The description of the domain	Open-ended
4	SIZE	Total size of the ontology	Open-ended
5	SIZEC	The total number of concepts	Open-ended
6	SIZEP	The total number of properties	Open-ended
7	SIZEA	The total number of axioms	Open-ended
6	SIZEI	The total number of fixed instances	Open-ended
7	LANG	The implementation language of the ontology	Scaled
8	METHOD	The ontology engineering methodology	Open-ended
9	PM	Ontology development effort in person months	Open-ended
10	SIZET	The size of the ontology engineering team	Open-ended
Part 2. Ontology engineering process			
11	SIZEB	Percentage of the final ontology built from scratch	Open-ended
12	DCPLX	Complexity of the domain analysis	Scaled
13	CCPLX	Complexity of the ontology conceptualization	Scaled
14	ICPLX	Complexity of the ontology implementation	Scaled
15	DATA	Complexity of the ontology instantiation	Scaled
16	SIZER	Percentage of the final ontology built by reuse	Open-ended
17	REUSED	Percentage of directly integrated reused ontology	Open-ended
18	REUSET	Percentage of integrated reused ontology after translation	Open-ended
19	REUSEM	Percentage of integrated reused ontology after modification	Open-ended
20	REUSETM	Percentage of integrated reused ontology after modification and translation	Open-ended
21	OU	Understanding of the reused ontologies	Scaled
22	OEREUSE	Complexity of the evaluation of the reused ontologies	Scaled
23	OT	Complexity of the translation of the reused ontologies	Scaled
24	OM	Complexity the modification of the reused ontologies	Scaled
25	OI	Complexity of the integration of different ontologies	Scaled
26	OETOTAL	Complexity of the evaluation of the final ontology	Scaled
28	REUSE	Required reusability of the ontology	Scaled
29	DOCU	Complexity of the documentation task	Scaled
Part 3. Engineering team			
30/31	OCAP/DECAP	Capability of the ontologists/domain experts	Scaled
32/33	OEXP/DEXP	Expertise of the team	Scaled
34/35	LEXP/TEXP	Level of experience with respect to languages and tools	Scaled
36	PCON	Personnel continuity	Scaled
Part 4. Software tool support			
37	TOOL	Level of technological support for particular ontology engineering activities	Scaled
38	SITE	Communication facilities in decentralized environments	Scaled

Table 1. The Questions Used in the Survey

ment projects, the ratio was 50%. We argue that this is a clear indicator for the level of maturity achieved by ontology engineering as a field at this point in time

- process-driven methodologies are used in one out of two projects in which assistance to the ontology engineering team is expected to be essential. A significant number of participants (approximately 20%) used METHONTOLOGY [Fernandez *et al.* (1997)] to develop their ontology. Just as in, for instance, software engineering, the fact that an increasing number of projects resort to predefined methodological support confirms the fact that ontology engineering is on its best way to become an established engineering discipline. This finding is novel compared to the ones of previous analytical and empirical surveys with similar objectives published two or more years ago. These surveys could not take into account the most recent, rapid advances of semantic technologies, and had a narrower scope in terms of the data set collected. As far as improvements are concerned, participants suggested that project settings in which domain analysis and evaluation needs run high mandate domain-specific customizations of the generic methodologies available. This can be confirmed by our data analysis, which indicates low tool support for these ontology engineering activities. High tool support therefore was shown to reduce development time considerably. Such customizations might be particularly beneficial for very complex domains, for the development of ontologies with broad coverage, or for those that involve non-common-sense knowledge such as life sciences. A last issue to be highlighted, in particular as more and more high-quality ontologies are becoming available, is ontology reuse. Our survey showed that this area of ontology engineering is still in an early stage of adoption. Conceptually, reuse is predominantly understood in a very broad sense, as capturing ontological knowledge from relevant information sources, including ontologies, but also other resources. When understood in its strict sense, it mainly consists of reusing a single ontology with minor adaptations of the original content (approximately 5%). The general scenario in which multiple reuse candidates are assessed and compared against each other is, independently of state-of-the-art of the research in the field, not covered by the projects surveyed. The same applies to the reuse of multiple ontologies which are customized, merged and integrated in a new application setting. While the adoption of ontology-based technologies will continue, it is likely that such scenarios will gain in relevance and more efforts will need to be invested in revising existing ontology reuse methods, techniques and tools towards providing the adequate level of support for non-technical users. A summary of the general issues is listed in Table 2.

4.2. *Process, Personnel and Project Issues*

As aforementioned, the structure of the survey assumed the activity breakdown of ontology engineering processes introduced in Section 2. This set-up proved to match to a satisfactory extent the way the surveyed projects carried out the process. The interviews emphasized, however, some discrepancies between

- the complexity of particular activities as perceived by ontology engineering practitioners,

Approximately 50% of the participants made use of an ontology engineering methodology in large-scale projects.
95% of the collected ontologies were built in Europe.
35% of the ontologies were built in the industry.
Approx. 60% of the ontologies were built from scratch.
If ontologies were reused in projects, they made up to 95% of the final ontology.
Ontology reuse was predominantly interpreted as usage of relevant information sources, be that ontologies or others. In other words, reuse is still mostly performed at a knowledge level as opposed to implementation level.

Table 2. Summary of General Issues

- the significance of these activities as measured in terms of their impact on the total development costs, and
- the level of maturity achieved at present by the R&D community with respect to methods and tools supporting these activities.

In the following we look into these issues in more detail. To investigate the relationship between the individual aspects, their interdependencies and impact on the development costs, we performed a correlation analysis. The *Correlation analysis* provided a general overview of the importance of each aspect, and assisted in identifying those aspects whose impact might have been underestimated so far and that would require additional attention in terms of novel methods and tools. Aspects can be positively correlated (value between 0.1 and 1), negatively correlated (value between -0.1 and -1) or independent (value between -0.1 and 0.1). Overall, the outcomes of the correlation analysis were consistent with the feedback we received from the interviewees. In some cases the values were counter-intuitive, which can be attributed to the high number of variables in the model and the diversity of projects involved in the data set.

4.2.1. *Correlation between Ontology Engineering Aspects and Effort*

Table 6 shows the correlation between the various ontology engineering-related aspects covered by the survey and the project effort in person months.

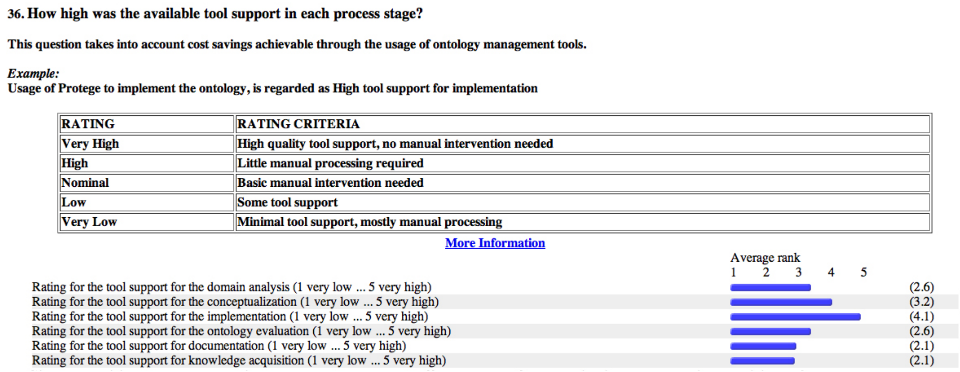
DCPLX Out of the six positively correlated factors, domain analysis was shown to have the highest impact on the total effort, achieving a significantly higher correlation value over the other five activities. This is an assessment of the time-consuming nature of the knowledge acquisition process, which was also confirmed by comments received from the participants in the interviews, and by previous surveys in the field. As the results in Figure 3 point out, tool support for this activity was very poor. Many interviewees questioned the utility of available tools, which were perceived as too generic

Ont. engineering 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	Percentage of integrated reused ontology	0.274
DOCU	Complexity of the documentation task	0.346
OE	Ontology evaluation	0.362
OCAP/DECAP	Capability of the ontologists/domain experts	-0.321
OEXP/DEEXP	Expertise of the ontologists/domain experts	-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

Table 3. Correlation between Ontology Engineering Aspects and Effort

especially when it came to ontologies developed for highly specialized domains such as health care, or in projects relying on end-user contributions. In addition, participants shared the view that process guidelines tailored for such specialized cases are essential for the success of ontology engineering projects. Current methodologies are very generic when it comes to issues of knowledge elicitation. They state the imperative need of a close interaction between domain experts and ontology engineers, but extensive studies on using techniques such as concept maps, card sorting and laddering [Cooke (1994)] are largely missing. These particular techniques, complemented with detailed insights on the practices established in the respective domains, could be very useful to design specially targeted methodologies and guidelines for ontology engineering.

OE The quality of the implemented ontologies remains a major concern among ontology engineers. Nevertheless, the projects we surveyed 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 the surveyed projects reported minor effort in formally testing the ontologies they developed. Other 48% reported fair use of simple testing methods which were carried out mostly manually. Only three 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. At this



36. How high was the available tool support in each process stage?

This question takes into account cost savings achievable through the usage of ontology management tools.

Example:
Usage of Protege to implement the ontology, is regarded as High tool support for implementation

RATING	RATING CRITERIA
Very High	High quality tool support, no manual intervention needed
High	Little manual processing required
Nominal	Basic manual intervention needed
Low	Some tool support
Very Low	Minimal tool support, mostly manual processing

More Information

	Average rank
Rating for the tool support for the domain analysis (1 very low ... 5 very high)	(2.6)
Rating for the tool support for the conceptualization (1 very low ... 5 very high)	(3.2)
Rating for the tool support for the implementation (1 very low ... 5 very high)	(4.1)
Rating for the tool support for the ontology evaluation (1 very low ... 5 very high)	(2.6)
Rating for the tool support for documentation (1 very low ... 5 very high)	(2.1)
Rating for the tool support for knowledge acquisition (1 very low ... 5 very high)	(2.1)

Fig. 3. Software Tool Support

juncture ontology evaluation plays a passive role for ontologies developed in less formal project settings such as in academia. However, as ontology evaluation practices increase with the demand for quality assurance, the associated impact on effort can be substantial.

DOCU Documentation proved to be a costly factor as well. The survey results point out that most of the developers of highly specialized ontologies perceived documentation as a resource-intensive activity. This was not necessarily true for less complex ontologies, or in cases in which the development process was less formal.

CCPLX, ICPLX The ontology conceptualization, which is responsible for the modeling of the application domain in terms of ontological primitives (concepts, relations, axioms), and the ontology implementation where the conceptual model is formalized in a knowledge representation language, are positively correlated factors. However, their impact on the total effort is not as high as the one of the domain analysis or the ontology evaluation. This outcome speaks for the relatively well understanding and high-quality tool support for these activities of the ontology engineering process.

OCAP/DECAP, OEXP/DEEXP, LEXP/TEXP The impact of personnel-related aspects suggests that more training programs in the area of ontology engineering, better collaboration support, and an improved, more fine granular documentation of the decisions taken during the ontology engineering process may have positive effects.

SITE The data analysis produced counter-intuitive results for the SITE parameter which accounts for the degree of distribution of the team and their communication and collaboration facilities. Here the analysis suggested that email communication lowered the effort needed to build ontologies while frequent face-to-face meetings increased the effort significantly. This could be based on the assumption that face-to-face meetings produced more different views on the ontology, and resulted in more discussions which, of course, raises the costs of ontology development.

The slight dominance of factors such as DCPLX (domain analysis) and OE (ontology evaluation) indicates that any facilitation in these activities may result in major efficiency gains. Even though tools such as wikis may be helpful especially in collaborative settings, they are still rarely used for the purpose of domain analysis. More generally, the results of the interviews indicated a low tool support for this task. This situation could be improved by applying methods such as automated document analysis, or ontology learning approaches, to support the analysis of the domain, the assessment of the information sources available and the knowledge elicitation process. Extending existing methodologies with specific empirically determined practices in place could also have a positive effect, particularly in vertical domains. A similar conclusion can be drawn for ontology evaluation. Despite of the availability of automated approaches such as unit tests, consistency checks or taxonomy cleaning techniques, ontology evaluation still lacks tools which are easy to use and comprehensible for most users.

Concluding the correlation analysis between ontology engineering aspects and effort, we can state that process activities such as domain analysis, conceptualization, implementation and evaluation, as well as the level of re-usability of the ontology, and the documentation requirements have a well-distributed correlation factor associated with the effort. This means that each of these activities exhibits a relevant impact on the effort, while at the same time indicating that no individual activity plays a overwhelmingly dominating role. As expected, the quality of the ontology engineering team is crucial for the success of a project; it would be interesting to investigate, however, the effect of such aspects in more collaborative scenarios, which could become the norm in ontology engineering. The data set on which this analysis is based on is not relevant for highly decentralized scenarios of community-driven ontology engineering. More research is needed to assess the state of the art in the area of ontology reuse and associated activities such as ontology understanding, merging, and integration. In this respect the survey is not representative and should be revisited once this engineering approach gains more importance, for instance, as a consequence of the wide scale development of a critical mass of ontologies in diverse vertical sectors.

4.2.2. *Correlation between Ontology Engineering Aspects*

In addition to the impact on the total development effort we analysed the correlation between specific aspects of ontology engineering projects. Since it is not possible to account for all possible relationships between them in the scope of this article, we will restrict ourselves to the most important findings in the following.

Personnel-related aspects (Table 4) were shown to be positively correlated. This was obvious for those questions referring to the capability and experience of the ontology engineering team. In most cases the survey showed that the capability of the participants was largely based on their project experience. Additionally, the software support available to projects carried out by the same ontology engineering team tended to remain unchanged. When new tools were introduced, the learning period for experienced practitioners was much higher than for novel developers. Similar observations were made in software engineering, in which habits of software use have a significant influence on acceptance and adoption of new software.

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

Table 4. Correlation between Personnel-Related Aspects

	DCPLX	**DOCU**
OE	0.211	0.389

Table 5. Correlation between Process-Related Aspects

High correlation values were also measured between activities within the ontology engineering process (Table 4). One, in particular, was between ontology evaluation and ontology documentation (Table 5). Data analysis showed that these results were largely concentrated on large-scale ontology engineering projects. This is possibly due to the fact that such ontology development projects run more extensive evaluation tests, which in turn might lead to additional documentation effort. Domain analysis was most highly correlated with the conceptualization and implementation. The majority of the interviewees did not perceive a clear cut between the conceptualization and the implementation activities. Conceptualization in most cases was a lightweight description and classification of the expected outcomes. In most of the projects surveyed there was no language- or tool-independent representation of the ontology. Instead, the ontology was implemented with the help of an ontology editor. In over 40% of the projects the development was performed mainly by domain experts, who generally agreed that current ontology editors are relatively easy to learn and utilize. This finding is different from the observations of previous surveys and comparative studies, and confirms one more time the fact that ontology engineering has reached an industry-strength level of maturity.

A summary of the results is presented in Table 6 below.[m]

[m] Average values refer to the following five-point scale 1:very low, 2:low, 3:nominal, 4:high, 5:very

Ont. eng. aspect	**Average value**	**Comments**
DCPLX	3.2	Knowledge acquisition support leaves room for improvement with respect to the level of detail of methodologies, availability of domain-specific best practices, guidelines and case studies, and usage of techniques such as card sorting or concept maps.
CCPLX	2.7	Conceptualization and implementation are well understood and supported through existing methodologies, methods and tools. In many cases there was no clear cut between the two activities and no language- or tool-independent representation of the ontology was produced.
ICPLX	2.4	
REUSE	2.7	Most ontologies are not subject to severe reusability requirements, thus the low assessment of this aspect.
DOCU	3.0	Developing highly specialized ontologies is associated to detailed documentation. This was not necessarily true for less complex ontologies, or in cases where the development process was less formal.
OE	2.0	Ontology evaluation still plays only a passive role in ontology engineering projects, thus the low average assessment of this aspect. However, when undertaken systematically, it accounts for a considerable share of the total costs of the project.
OCAP/ DECAP	3.8	Personnel-related aspects were crucial for the success of the surveyed projects, particularly for classical ontology engineering. The personnel turnover was not relevant for most of these projects, thus the comparatively low value assigned. As collaborative ontology engineering gains more adoption, it would be interesting to identify which aspects still remain relevant for the total development costs and which ones need to be added to account for a community-driven approach (e.g., motivation). The capability of the contributors largely correlates with their experience in related projects.
OEXP/ DEEXP	3.6	
PCON	1.8	
LEXP/ TEXP	3.3	
SITE	3.4	This accounts for the overhead associated with interactions within a geographically distributed project team. Face-to-face meetings were perceived to increase the total costs of the project, an issue which demands for adequate collaborative environments which support decentralized ontology engineering.

Table 6. Summary of Process, Project, and Personnel Issues

5. Related Work

This section gives an overview of the surveys and case studies on ontology engineering previously published in the literature. With respect to the surveys we can distinguish between analytical and empirical ones. The first category of surveys analyses ontology engineering methodologies from the theoretical perspective and identifies open issues in this field. In the nineties the focus was on laying out the foundations of ontology engineering by defining representation languages, designing methodologies supporting the ontology life cycle, proposing modeling

high.

patterns, and developing ontologies for particular domains [Peterson *et al.* (1998); Lenat *et al.* (1995); Noy *et al.* (1997)]. Surveys from that period mention, for instance, the need for guidelines and best practices on ontology development and reuse, and for fully fledged, tool-supported methodologies [Jones *et al.* (1998); Grüninger *et al.* (1995)]. Some of these issues were still to be solved years later: a comprehensive survey on ontology engineering methodologies published in 2003 identified a lack of software support for many methodologies and their limited scope as compared to the ontology life cycle [Corcho *et al.* (2003)]. More recent surveys highlighted the integration of ontology engineering into the enterprise (business process) modeling landscape and the study of ontology engineering economics as essential issues for the adoption of ontology engineering beyond the research labs of the Semantic Web community [Sure *et al.* (2006)]. A second category of surveys focuses on insights and findings derived from real-world case studies on ontology development, management, or reuse. Most of the existing surveys reported on the application of self-developed methodologies, highlighting their advantages as compared to alternative engineering approaches [Uschold *et al.* (1998); Tempich *et al.* (2006); Uschold *et al.* (1995)]. A common theme is the need for a software environment supporting the ontology engineering team at particular stages of the methodology, as a means to reduce the associated costs and to lower the barrier of entry for potential applicants. Other surveys described the deployment and evaluation of a specific ontology engineering methodology in a specific domain or system, or introduced in detail a particular ontology engineering tool and its usage in a given context [Bernaras *et al.* (1996); Gangemi *et al.* (1998); Lau *et al.* (2002)].

Last, but not least, there are surveys reporting on practical experiences in ontology-related projects [Paslaru *et al.* (2005); Niemann *et al.* (2006)]. Case studies such as [Hristozova *et al.* (2003); Koenderink *et al.* (2005); Tautz *et al.* (2000)] evaluated ontology engineering technology with respect to their relevance and usability, prior to applying them in a particular application setting, or operated the engineering process without nominally committing to existing techniques. The case studies generally assessed the limited usability or the poor impact of most parts of existing ontology engineering methodologies, methods and techniques. Practical guidelines and recommendations for developing ontology-based applications in specific sectors are available, for instance, in [Mochol *et al.* (2006); Noy *et al.* (2001)].

Other studies highlighted the limited awareness of ontology engineering methodologies in commercial settings: [Cardoso (2007)] analysed key trends and developments in a comprehensive survey which involved several hundreds of Semantic Web researchers and practitioners. The survey reported on the scope of ontology development projects, as well as languages, methodologies, and tools in use, but had a broader scope and thus provided fewer insights on ontology engineering practice. One of their conclusions was the limited awareness of ontology engineering methodologies among practitioners which, authors argued, hampered industrial adoption. Another study reported on social and technical bottlenecks which hinder the wide uptake of ontologies and one of its main findings was the need for advanced technology to cope with ontology development and maintenance especially in rapidly changing domains [Hepp (2007)].

6. Conclusions

Industry is starting to acknowledge the technical value of ontologies for enterprises. In the last years, early adopters have been increasingly using them in various application settings ranging from content management to enterprise application integration. The main technological building blocks and development platforms are meanwhile available from established vendors. Despite this promising position, the information known about the process underlying the development of ontologies in practice is still very limited. The literature reports predominately on case studies which involved methodologists, while ontologies are envisioned to be developed by domain experts possessing limited to no professional skills in ontology engineering. Surveys of recent date, including previous work of ours from 2006, are either of analytical nature or not supported by a critical mass of the community of adopters.

The aim of this paper was to fill this information gap through a study with 148 projects that developed ontologies for commercial as well as academic applications for a wide range of domains. This is by far the largest survey of this kind conducted in the community, and covers a significant share of the most relevant and popular ontology engineering projects ever run in the rapidly evolving semantic technology landscape.

The survey investigated the systematics, the development effort and the problems encountered in ontology engineering projects. We collected answers to 38 questions related to common phases of ontology engineering, as well as to personnel and project setting. The main findings of the survey are:

- ontology engineering methodologies are used in critical projects developing large ontologies or being under critical requirements;
- an increasing number of ontology projects involve end-users in the development projects. Ontology editing tools seem to be well-appropriate to be used by less technology-prone users, whilst project teams are reluctant to changing a tool environment due to the high learning curve;
- ontology engineers need cost benefit analysis methods to determine the transition point between ontology engineering activities; and
- with the uptake of ontology-based technology the need arises for methodologies and techniques customized for the characteristics of particular vertical domains. This is most notably true for activities having a high impact on the total development costs such as domain analysis and evaluation.

These findings confirm the fact that ontology engineering 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.

In order to overcome some of the problems discovered though this survey we suggest to:

- put some effort in adapting and refining generic ontology-based technology for settings in which the domain analysis and the evaluation of the resulting ontology are acknowledged to be challenging. Ontology engineering research should investigate practices and techniques already in place in the respective communities and integrate these in domain-specific versions of methodologies and software tools, most notably for knowledge acquisition and evaluation;
- provide support for a detailed documentation of decisions taken during the ontology engineering process;
- investigate in more comprehensive studies the state of the art in ontology reuse and decentralized ontology engineering;
- set up specific targeted training programs for practitioners and end-users; and
- design instruments for assessing the costs and benefits of ontology engineering.

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

Astrova, I. (2004). *Reverse engineering of relational databases to ontologies*, In *Proceedings of the First European Semantic Web Symposium ESWS2004*, 327–341.

Benjamin, P. C., et al. (1994). *Ontology capture method (IDEF5)*, Technical report, Knowledge Based Systems Inc.

Bernaras, A., Laresgoiti, I., Corera, J. (1996). *Building and reusing ontologies for electrical network applications*, In *European Conference on Artificial Intelligence (ECAI’96).*

Braun, S., Schmidt, A., Walter, A., Nagyp, G., Zacharias, V. (2007). *Ontology Maturing: a Collaborative Web 2.0 Approach to Ontology Engineering*, In Proceedings of the Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007) at the 16th International World Wide Web Conference (WWW2007).

Cardoso, J., (2007). The Semantic Web Vision: Where Are We?. *IEEE Intelligent Systems*, **22(5)**:84–88.

Cimiano, P. (2006). *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications.* Springer.

Cooke, N. J.(1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies*, **41**:801–849.

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.

Dimitrova, V., Denaux, R., Hart, G., Dolbear, C., Holt, I., Cohn, A.G. (2008). *Involving Domain Experts in Authoring OWL Ontologies.* In Proceedings of the 7th International Semantic Web Conference (ISWC2008), 1–16.

Euzenat, J., Shvaiko P. (2007). *Ontology Matching.* Springer.

Fernandez, M., Gomez-Perez, A., Juristo, N. (1997) *Methontology: From ontological art towards ontological engineering.* In Proceedings of the AAAI'97 Spring Symposium on Ontological Engineering.

Gangemi, A., Pisanelli, D., Steve, G. (1998). Ontology integration: Experiences with medical terminologies. *Formal Ontology in Information Systems*, 163–178, IOS Press.

Gómez-Pérez, A., Fernández-López, M., Corcho O. (2003). *Ontological Engineering.* Springer.

Grninger M., Fox M. (1995). *Methodology for the design and evaluation of ontologies.* In Proceedings of the IJCAI'95, Workshop on Basic Ontological Issues in Knowledge Sharing.

Hepp M. (2007). Possible ontologies: How reality constrains the development of relevant ontologies. *IEEE Internet Computing*, **11(1)**:90–96.

Hepp, M., de Bruijn J. (2007). Gentax: A generic methodology for deriving owl and rdf-s ontologies from hierarchical classifications, thesauri, and inconsistent taxonomies. In E. Franconi, M. Kifer, and W. May (editors) *Lecture Notes in Computer Science*, **4519**:129–144. Springer.

Hristozova, M., Sterling, L. (2003). *Experiences with Ontology Development for Value-Added Publishing.* In Proceedings of the International Workshop on Ontologies in Agent Systems OAS03 co-located with the AAMAS.

Jones, D., Bench-Capon, T., Visser, P. (1998). Methodologies for ontology development. *IT & KNOWS Conference of the 15th IFIP World Computer Congress*, 62–75.

Koenderink, N. J. J. P., Top, J. L., van Vliet, L. J. (2005). *Expert-based ontology construction: A case-study in horticulture.* In Proceedings of the International Database and Expert Systems Applications Workshops (DEXA05).

Kotis, K., Vouros, G. A. (2005). Human-centered ontology engineering: The HCOME methodology. *Knowledge and Information Systems*, **10(1)**:109–131.

Lau, T., Sure Y. (2002). *Introducing ontology-based skills management at a large insurance company.* In Proceedings of the Modellierung 2002, 123–134.

Lenat, D. (1995). Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, **38(11)**:33–38.

Maedche, A. (2002). *Ontology Learning for the Semantic Web*. Kluwer Academics.

Mainz, I., Weller, K., Paulsen, I., Mainz, D., Kohl, J., von Haeseler, A. (2008). Ontoverse: Collaborative Ontology Engineering for the Life Sciences. *Information Wissenschaft & Praxis*, **2**:91–99.

Mochol, M., Simperl, E. (2006). *Practical guidelines for building semantic eRecruitment applications.* In Proceedings of International Conference on Knowledge Management (iKnow'06), Special Track: Advanced Semantic Technologies (AST2006).

Niemann, M., Mochol, M., Tolksdorf, R. (2006). *Improving online hotel search - what do we need semantics for?.* In Proceedings of Semantics 2006 (Application Paper).

Noy, N. F., Hafner, C.D. (1997). The state of the art in ontology design: A survey and comparative review. *AI Magazine*, **18(3)**:53–74.

Noy, N. F., McGuinness, D.L. (2001). *Ontology development 101: A guide to creating your first ontology.* Technical Report Stanford Knowledge Systems Laboratory Technical, Report KSL-01-05 and Stanford Medical Informatics Technical Report, SMI-2001-0880, Stanford University.

Paslaru-Bontas, E., Mochol, M., Tolksdorf, R. (2005). *Case studies on ontology reuse.* In Proceedings of the 5th International Conference on Knowledge Management (I-Know'05).

Paslaru-Bontas, E. (2007). *A Contextual Approach to Ontology Reuse: Methodology, Methods and Tools for the Semantic Web.* PhD thesis, Freie Universität Berlin.

Paslaru-Bontas, E., Tempich, C. (2006). *Ontology Engineering: A Reality Check.* In Proceedings of the 5th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE2006), 836–854, Springer.

Peterson, B. J., Andersen, W. A., Engel, J. (1998). *Knowledge bus: Generating application-focused databases from large ontologies.* In Proceedings of the 5th KRDB Workshop Seattle.

Pinto, H. S., Martins, J. (2000). *Reusing ontologies.* In AAAI 2000 Spring Symposium on Bringing Knowledge to Business Processes, 77–84.

Pinto, H. S., Tempich, C., Staab, S. (2004). *Diligent: Towards a fine-grained methodology for distributed, loosely-controlled and evolving engingeering of ontologies.* In R. L. de Mantaras and L. Saitta, (editors) Proceedings of the 16th European Conference on Artificial Intelligence (ECAI 2004), 393–397, IOS Press.

Simperl, E., Tempich, C. (2007). A Methodology for Ontology Learning. In P. Buitelaar and P. Cimiano (editors) *Bridging the Gap between Text and Knowledge - Selected Contributions to Ontology Learning and Population from Text*, IOS Press.

Siorpaes, K., Hepp, M. (2008). Games with a purpose for the semantic web. *IEEE Intelligent Systems*, **23(3)**:50–60.

Sure, Y., Staab, S., Studer, R. (2002). Methodology for development and employment of ontology based knowledge management applications. *SIGMOD Record*, **31(4)**:18–23.

Sure, Y., Tempich, C., Vrandecic, D. (2006). Ontology Engineering Methodologies. *Semantic Web Technologies: Trends and Research in Ontology-based Systems*, 171–187, Wiley.

Tautz, C., Althoff, K. D. (2000). *A case study on engineering ontologies and related processes for sharing software engineering experience.* In Proceedings of the International Conference on Software Engineering and Knowledge Engineering (SEKE00).

Tempich, C., Pinto, H.S., Staab, S. (2006). Ontology engineering revisited: an iterative case study with diligent. In Y. Sure and J. Domingue (editors) *The Semantic Web: Research and Applications: 3rd European Semantic Web Conference(ESWC 2006)*, **4011**:110–124, LNCS, Springer.

Tempich, C., Simperl, E., Pinto, S., Luczak, M., Studer, R. (2007). Argumentation-based Ontology Engineering. *IEEE Intelligent Systems*, **22(6)**:52–59.

Uschold, M., King, M. (1995). *Towards a methodology for building ontologies.* In Workshop on Basic Ontological Issues in Knowledge Sharing.

Uschold, M., King, M., Moralee, S., Zorgios, Y. (1998). The enterprise ontology. *Knowledge Engineering Review*, **13(1)**:31–89.

Van Damme, C., Hepp, M., Siorpaes, K. (2007). *Folksontology: An integrated approach for turning folksonomies into ontologies.* In Bridging the Gep between Semantic Web and Web 2.0 (SemNet2007), 57–70.