A SYSTEMIC METHODOLOGY FOR ONTOLOGY LEARNING
An Academic Case Study and Evaluation

Richard Gil, Leonardo Contreras
Dept. of Processes and Systems, Simón Bolívar University, Caracas-Baruta, Venezuela

María J. Martín-Bautista
Dept. of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain

Keywords: Ontology Learning, Methodology, Systemic, Methodology Evaluation, Tools, Academic Domain.

Abstract: There is an important dispersion of technical and methodological resources to support the complete Ontology Learning (OL) process from diverse knowledge sources. This fact makes the maintaining of the structures of representation (ontologies) difficult. Therefore, the Knowledge-based Systems associated with user's domains may not fulfil the increasing knowledge requirement from the user. In this paper, we give a possible solution for this problem. For this purpose, we propose a Systemic Methodology for OL (SMOL) that unifies and simplifies to the users the whole process of OL from different knowledge sources (ontologies, texts and databases). SMOL as methodology is evaluated under DESMET methods, in addition with their application for an academic case study is also included.

1 INTRODUCTION

Reaching the knowledge from a semantic technological perspective has propitiated the development of new technical methodologies and resources. Through these product-tools obtained the society should explore, discover, recover, store and update knowledge associated to some specific domains (Decker et al., 2000)(Nonaka, 1994).

The aim is not only to obtain system products to support user requirements that may be 'devalued' throughout time, but also to reach Knowledge-based System (KBS) able to auto-learn and to make recommendations and learning actions related to different user communities (Borges et al., 2008) (Garruzzo et al., 2007).

However, this kind of systems is not so easy to develop and maintaining (Abdullah et al., 2006). This is due several aspects: first, the ontology engineering methodological resources are in maturation process yet; second, even though technologies for handling knowledge based on ontologies satisfy some user requirements, they cannot guarantee a complete quality-driven and user-oriented development during the ontology engineering process; and finally, partial experiences of ontology engineering developers, researchers and users may be not incorporated as part of the tacit knowledge (behaviour and skills) in the new methodologies and technologies yet (Gómez-Pérez and Manzano-Macho, 2005) (Haase et al., 2005).

Despite there are various definitions about OL, we are in according to (Gómez-Pérez and Manzano-Macho, 2005), where it is "*the application of a set of methods and techni-ques used for improving a previous ontology with heterogeneous Knowledge Sources (KSo), avoiding the complete Ontology Development process*". These sources can be ontologies previously developed, texts, database or results of a process of ontology integration (Maedche and Staab, 2001).

Precisely the widespread variety of mechanisms and resources for OL make difficult the definition of a standard methodology for OL. Consequently, in (Gil, 2009) a new Systemic Methodology for OL (SMOL) is conceived and proposed to overcome some identified restrictions.

In this paper, we focus on the OL processes in Section 2. The systemic methodology perspective is showed in Section 3. SMOL methodology is descri-

Gil R., Contreras L. and J. Martín-Bautista M..
A SYSTEMIC METHODOLOGY FOR ONTOLOGY LEARNING - An Academic Case Study and Evaluation.
DOI: 10.5220/0003070602060212
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* (KEOD-2010), pages 206-212
ISBN: 978-989-8425-29-4

bed in Section 4. A case study in the academic domain applying SMOL is included in Section 5. A general methodology evaluation is applied to SMOL in Section 6, and finally, conclusions in Section 7.

2 ONTOLOGY LEARNING

There are several methodological alternatives in the literature about OL. The one suggested in (Maedche and Staab, 2001) includes some learning approaches besides possible and recommended set of activities associated with them. On the other hand, in (Gliozzo et al., 2007), a different classification of the recommended techniques into two groups are given. The first group includes those approaches that allow getting knowledge and retrieving information from electronic texts. The second group, includes those approaches that allow to 'gain knowledge' based on previous structured knowledge and ontologies such as dictionaries and thesaurus (Gómez-Pérez and Manzano-Macho, 2005)(Gacitua et al., 2008).

OL approaches according with the KSo are three-fold: a) OL from other ontologies developed previously (Ehrig, 2007)(Noy and Musen, 2000) (Euzenat et al., 2007) b) OL from documents (Buitelaar and Cimiano, 2008) (Cimiano, 2006). And c) OL from database schemes and their data-values (Astrova et al., 2007)(Nyulas et al., 2007) (Cerbah, 2009).

2.1 Ontology Learning Resources

There are some definitions regarding Methodological Resources (MR) that allows us to understand the concepts associated to MR and to avoid confusions that sometimes happen in technical literature. The following definitions (Callaos, 1992) have been considered: a) *Techniques*: subjective capabilities (abilities or skills) to handle a tool properly. b) *Methods*: a way of thinking or doing using a tool to achieve an objective. c) *Tools*: objective capabilities to use the resources properly to apply techniques. And, d) *Methodologies*: a related set of methods, techniques and tools which could be used for reaching objectives.

2.2 Ontology Learning Problematic

Although important technical advances in MR in the OL field, according to each KSo, have demonstrated the main OL strengths and opportunities, authors recently have reported high dispersion and little integration among those MR producing OL results from the same KSo.

Therefore, to synthesize the general OL problems a situational technical analysis, which is known as SWOT (Strengths, Weaknesses, Opportunities and Threats), has been used (Hill and Westbrook, 1997). This technique simplifies the OL understanding from two broad perspectives. First, it addresses the knowledge development and reconstruction as an OL process and, secondly, by studying it in terms of the resulting semantic (Noy and Klein, 2004).

In agreement with (Gómez-Pérez and Manzano-Macho, 2005) and (Shamsfard and Abdollahzadeh, 2003) some conclusions associated with those studies about OL methodologies can be summarized.

Regarding OL Methods: a) There is not an established standard. b) The methods are not usually combined, and c) Many methods are not associated with specific tools. With regard to OL Tools: a) All of them help to extract knowledge; b) A small group of them allow the retrieval of a complete taxonomy, and c) Only some tools support specific OL methods.

It is possible to infer also, that OL methodological options do not exist as a complete integral, unified and dynamic way to face the OL problems for knowledge recovery from heterogeneous KSo.

3 SYSTEMIC PERSPECTIVE

Methodological options used to get designs and knowledge product development (systems, models or ontologies) are associated to strategies and processes structured in some way. Many approaches closer to Software Engineering reflect the efforts dedicated in this direction (Sommerville, 2006).

More specific methodological approaches orient-ted on one hand to Software Development (Press-man, 2006) and on the other hand, to Knowledge Engineering (Gómez-Pérez et al., 2004) (Buitelaar and Cimiano, 2008)(De-Nicola et al., 2009) the previous methodological options have arisen.

The proposed methodological perspective tries to conciliate the system development total quality paradigms with user-centered services to attend their demanded requirements. This conciliation is supported by systemic methodologies instead of systematic ones (Callaos and Callaos,2003): a) *Systematic methodologies* are oriented to the efficiency, with a predetermined behaviour, strict and closed. (e.g. Structured Life Cycle) and b) *Systemic methodologies* are oriented to the effectiveness, with a non-

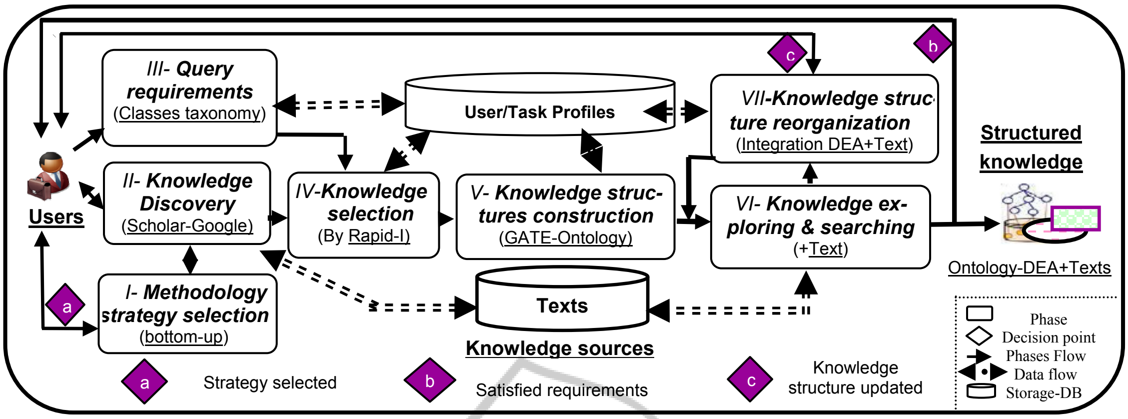

Figure 1: Systemic Methodology for Ontology Learning applied for an academic case study.

predetermined behaviour, flexible and open. (e.g. Agile Process).

Applying a methodology to an environment (an organization or system) is an evolutionary maturing process to achieve results with approaches such as the Action-Research (Baskerville, 1999) and the Action-Learning (Dilworth, 1998). The action (Action-Design) in both cases allow to support: first, researching to discover; second, learning to understand and to experiment; and third, system design and synthesis to generate new ideas and to solve specific problems to satisfy certain requirements.

Indeed, the product (ontologies) and the process (methodologies) must be developed in a trade-off between efficient and effective action-design.

4 DESCRIPTION OF SMOL

The lack of integrated methodologies covering the whole process of OL leads us to propose and experiment with new methodological options. This new systemic methodological proposal must be flexible, iterative, incremental and adaptable to normal users, experts and knowledge engineers using some MR previously developed according the quality approach cited (Callaos and Callaos, 2006).

Users of SMOL to combine MR for diverse KSo in a proper way, considering the existence of a domain ontology already elaborated for KBS which could be improved through updating/enrichment OL processes (Haase et al., 2005)(Noy and Klein, 2004).

For this methodology design, we select a framework for knowledge retrieval of (Yao et al., 2007).

4.1 SMOL Phase-flow Description

The phase-flow of SMOL is proposed (some in Figure 1), emphasizing the MR recommended to be used in each specific phase. The activities related to each original phase of SMOL are explained as follow: *I. Methodology strategy selection*. The complexity of the domain is evaluated based on the availability of information/knowledge useful about the domain (Zhou, 2007). The methodology strategy is drafted/selected using an appropriated arrangement of MR for each KSo relative to. *II. Knowledge discovery.* The MR from different knowledge -sources and -repositories are combined. *III. Query requirements.* Different queries are formulated to the KSo available by browsers or other kind of applications. *IV. Knowledge selection*. A selection of the retrieved information from the formulated queries to the sources and repositories is performed. *V. Knowledge structures construction.* Different structures such as ontologies and contexts can be built interacttively with users' advisory by ontology alignment, machine learning techniques, etc. *VI. Knowledge exploring and searching.* The knowledge structures are explored, verified and validated and the search can be refined. *VII. Knowledge structures reorganization.* Processes such as grouping of instances, ontology population and other similar activities are performed in this phase. And, *VIII. Knowledge-based System configuration*. Users set-up the main modules of the KBS that have ontologies updated and associated with the users' domain.

Other five activities were developed for SMOL drafting: 1) Methodology strategic selection phase is designed considering that the user of SMOL may adjust MR according to the information available in the KSo about the domain-complexity. 2) Knowledge Sources are configured as storage component (DB) with the purpose of knowledge reusing. 3) User/Task profiles are configured as storage components to queries-operations registration (log) with the purpose of reusing MR and recommending tasks. 4) Decision points have been included for user cyclical-quality-check purpose. Some of them are shown –as rhombus- in the Figure 1. And, 5) A methodological Phase-flow activities description is detailed as input, output, methods and tools recommended according with each KSo and strategy selected.

An ontology graphical tool option for ontology management was used in GATE to display annotations to the users and to help them to support ontology updating (evolution) from texts (Phases V &VII).

Other SMOL applications for the same case study using other KSo have been reported: a) OL by comparing to domain ontology located and recovered from the Internet (Gil et al., 2008). And, b) relevant knowledge about profiles of professor's subdomain from a relational database (RDB) of another University was obtained (Gil et al., 2010).

5 ACADEMIC CASE STUDY

The objective of this study in the academic domain is to retrieve and to add new knowledge into an ontology named Ontology-DEA previously developed for a Decision-Support System of a University of Venezuela (Ramos and Gil, 2007). In this experimental case, it was improved by users the ontology-DEA (incremental/iterative) with knowledge extracted from a corpus of texts. SMOL application for this case is shown in Figure 1 (Gil et al., 2009).

A bottom-up learning strategy (Phase I) was draf-ted and selected considering the following key acti-vities: a) Finding and selecting a set of texts from Internet with experts-users advisory. b) Identifying from the corpus, some relevant keywords by agent for ontology updating. And, c) applying the OL from texts via text annotations and ontology population.

Texts selection with user's participation is carried out in these Phases (II & III). Users recovered texts for the corpus, through Google Scholar. From an initial set of 1000 retrieved texts, a final set of 480 texts were selected using a file-length base.

The learning agent developed in RAPID-I with the plug-in WVtool was used to classify texts by their relevant keywords, so they could be added to the corpus for future updating (Phases III & IV). The used technique is "text clustering" with the TF-IDF term weighting scheme. Moreover, different processses of tokenization, stop-word removing and stemming were performed. Keywords found by the agent were: "accredit, style, programming, distance, institute, program, online, faculties, course and student". Those Keywords selected by the agent were inputs to the next process in GATE via Onto Gazetteers (Phase VI). The central purpose was to identify representative terms and concepts in the texts of the corpus besides corresponding Gazetteers' annotation standard (e.g. dates or places).

6 SMOL EVALUATION

There are not so many alternatives for methodology evaluations applied to the Ontology Development field. One of the most referred in the Software Engineering area is DESMET (Kitchenham, 1996).

We have used a combination of these DESMET methods: a) *Screening:* A feature-based evaluation done by a single individual who not only determines the features to be assessed and their rating scale but also does the assessment. For initial screening, the evaluations are usually based on literature describing the software method/tools rather than the actual use of the methods/tools. b) *Experiment:* A feature-based evaluation done by a group of potential users who are expected to try out the methods/tools on typical tasks before making their evaluations. And, c) *Case study:* A feature-based evaluation performed by someone who has used the method/tool on a real project.

Those methods are recommended by DESMET to be used when: a) Large number of methods/tools to assess. b) Short timescales for evaluation exercise. c) Benefits difficult to quantify. d) Benefits observable on a single project. e) Stable development procedu-res. f) Relatively small learning time. And, g) Tool-/method user population very varied and limited.

6.1 Qualitative Screening

To apply the Qualitative screening of the DESMET evaluation, we have followed the two-fold: First, we have performed below an interesting evaluation approach applying usability/suitability criteria assessment to evaluate by comparing our proposal (Dahlem and Hahn, 2009).

On the other hand, we have developed a short comparison of SMOL among two similar OL methodologies published recently by (Simperl et al., 2008) and (Novacek et al., 2007).

6.1.1 Screening through Usability Criteria

Dahlem's usability and suitability evaluation proposal (Dahlem and Hahn, 2009) has considered the thirteenth methodological criteria suggested and applied. Those are: Adequate terminology (C1); Structure (C2); Descriptiveness (C3); Transparency (C4); Error avoidance (C5); Robustness (C6); Lookahead (C7); Consistency (C8); Hiding formality (C9); Expressiveness (C10); Conceptualization flexibility (C11). Ontology assumptions (C12); and, Tool support (C13). These criteria are combined originally in an upper level under the following terms: I) Learnable. II) Efficiency. III) Memorability. IV) Error Handling. And, V) Satisfaction.

First, we assessed the SMOL methodology characteristics with the usability criteria cited. As result, SMOL has up to nine of thirteen representative criteria for methodology usability according to Dahlem's proposal. According to the total usability evaluation criteria (uniform presence considered), Efficiency and Satisfaction for SMOL, the value is high (0,85 & 0,75). As for the Learnable, Memorability and Error-handling criteria about SMOL are medium (0,5). Indeed, these methodology evaluation results about SMOL show comparative feature-/capabilities among other equivalent methodologies.

6.1.2 Screening through Comparison

Comparing the SMOL methodology among the proposal of Simperl et al. (Simperl et al., 2008) and DINO (Novacek et al., 2007), there are some aspects to point out. Mainly, SMOL has more elaborated methodological options to support OL processes from diverse KSo. The main options are: 1) SMOL considers explicitly the assessment of the domain-complexity characteristics for strategy selection. And, 2) the OL strategy selection is based on an approach of learning to start not always from texts, but from other KSo such as databases and ontologies, inclusive at beginning of the process. Some details are omitted due paper-pages limit.

6.2 Qualitative Experiment And Case

The main way to test the SMOL functionality is based on the case study, because we can check the user validation and experiment with related methods/tools. Some users were trained/familiarized with some OL methods/tools used (e.g. Protégé, Prompt-CogZ, Racer-Pro, GATE).

For each KSo, an evaluation strategy has been designed considering: context (goals-constrains), planning and design, preparation, execution, data analysis, dissemination and decision-making.

An interview-questioner was given to the users (up to 6) during the OL cycle according to each KSo. A feature-based analysis was applied to those results associated with those MR used. Particularly, we asked them about Tools-functionality and Input-/Output related to the OL methods/techniques learned/applied.

The case study evaluation revealed the user satisfaction about the SMOL methodology flexibility, due to the capability of the MR integration in the systemic component. A minor issue very interesting for the user is the diversification of strategies to reach knowledge aggregation from different KSo.

7 CONCLUSIONS

There is a lack of integrated methodologies in the OL process, whatever the sources considered: ontologies, texts and database. A Systemic Methodology for OL named SMOL has been designed considering pros and drawbacks of the previous OL methodolgical proposals but including MR for diverse KSo. The result is an integral, flexible, open, interactive and iterative methodology user-oriented.

The SMOL methodology has been applied and checked in an academic case study for different KSo. Particularly, the OL from texts has been detailed in this work.

The ontology updated by user's participation help us for SMOL validation. The SMOL methodology evaluation as well as the preliminary result obtained for this case study, reveal the feasibility of SMOL as an instance of new methodological perspectives for OL from texts, as a way to update ontologies asso-ciated with KBSs of the users' domain.

In future works, promissory results could be obtained with other SMOL cases applications combining incrementally some different KSo. Likewise, other experimental and specific evaluations have been performed to increase the SMOL background.

REFERENCES

Abdullah, M., Kimble, C., Benest, I., and Paige, R. (2006). *Knowledge-based systems: a re-evaluation. Journal of Knowledge Management*, 10 Nro 3:127–142.

Astrova, I., Korda, N., and Kalja, A. (2007). *Rule-based transformation of sql relational databases to owl* *ontologies.* Proceedings of the 2nd International Conference on Metadata & Semantics Research.

Baskerville, R. (1999). *Investigating information systems with action research.* Comm.AIS, v:2: Art 19.

Borges, A., Corniell, M., Gil, R., Contreras, L., and Borges, R. (2008). *Towards a study opportunities recommender system in ontological principles-based on semantic web environment.* In The 4th WSEAS/IASME. (EDUTE'08), ACM vol 8, no 2.

Buitelaar, P. and Cimiano, P. (2008). *Ontology Learning and Population: Briging the Gap Between Text And Knowledge*. IOS Press, Netherland.

Callaos, N. (1992). *A systemic system methodology*. In International Conference on System Research Informatic and Cybernetics, Baden-Baden, Germany.

Callaos, N. and Callaos, B. (2003). *Toward a practical general system methodological theory*. Journal of Systemics, Cybernetics and Informatics, 1:114–120.

Callaos, N. and Callaos, B. (2006). *Designing with a system total quality*. In on Information System Analysis, and Synthesis, ISAS'06, USA, p 15–23.

Cerbah, F. (2009). *RDBToOnto User Guide, Version 1.2 Beta From relational Database to Fine-Tuned Populated Ontologies.* http://www.tao-project.eu/

Cimiano, P. (2006). *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications.* Springer-Verlag New York, LLC.

Dahlem, N. and Hahn, A. (2009). *User-friendly ontology creation methodologies-a survey.* 15th Amer. Conf. On Information Systems, California-USA, pages 1–9.

De-Nicola, A., Missikoff, M., and Navigli, R. (2009). *A software engineering approach to ontology building. Information Systems*, Elsevier, 34:258–275.

Decker, S., Melnik, S., Van-Harmelen, F., Klein, D., Fensel, F., Broekstra, M., Erdmann, J., and Horrocks, M. (2000). *Knowledge networking the semantic web: The roles of xml and rdf.* IEEE Internet Computing.

Dilworth, R. (1998). *Action learning in a nutshell.* Performance Improvement Quarterly, Vol 11(1).

Ehrig, M. (2007). *Ontology Alignment: Biding the Semantic Gap.* Book, Springer-Verlag.

Euzenat, J., Mocan, A., and Sharffe, F. (2007). *Ontology Management, Ontology Alignments: An Ontology Management Perspective*. Book, Springer.

Gacitua, R., Sawyer, P., and Rayson, P. (2008*). A flexible framework to experiment with ontology learning techniques.* Know.-Based Syst., 21(3):192-199.

Garruzzo, S., Rosaci, D., and Sarné, G. (2007). *Mars: An agent-based recommender system for the semantic web*. LNCS, 4531:181–194.

Gil, R. (2009). *New systemic methodology framework for ontology learning (in spanish).* Master's thesis, Dpto.Computer Science, Granada University, Spain.

Gil, R., A.Borges, Ramos, L., and Contreras, L. (2008). *Ontologies integration for university institutions: Approach to an alignment evaluation*. In Proc. 19th Australian Conference on Software Engineering ASWEC 2008, pages 570–578.

Gil, R., Borges, A., Contreras, L., and Martín-Bautista, M. (2009). *Improving ontologies through ontology learning: a university case*. In CSIE'09, IEEE Computer Society, March-April, L.A.-USA.

Gil, R., Martín-Bautista, M., and Contreras, L. (2010). *Applying an ontology learning methodology to a relational database: University case study.* IEEE-ICSC-2010, USA, Sep 22-24 (Accepted).

Gliozzo, A., C. Caracciolo, M. D-Aquin, M. S., Peter, W., Voelker, J., Dzbor, M., Mota, E., Gomez-Perez, A., Haase, P., Waterfield, W., Contreras, J., Grobelink, M., Euzenat, J., Cunning, H., Staab, S., Gangemi, A., Angele, J., Iglesias, M., Lobo, T., and Lopez, A. (2007*). Results from experiments in ontology learning including evaluation and recommendation*. Technical Report, http://www.neon-project.org/

Gómez-Pérez, A., Fernando-López, M., and Corcho, O. (2004). *Ontology Engineering.* Book, Springer-Verlag, London- UK.

Gómez-Pérez, A. and Manzano-Macho, D. (2005). *An overview of methods and tools for ontology learning from text.* Knowledge Engineer. Rev., 19:187–212.

Haase, P., Volker, J., and Sure, Y. (2005). *Management of dynamic knowledge.* Journal of Knowledge Management, 9:97–107.

Hill, T. and Westbrook, R. (1997). *Swot analysis: It's time for a product recall.* Long Range Planning, 30:46–52.

Kitchenham, B. (1996). *Evaluating software engineering methods and tool. part 3: Selecting an appropriate evaluation method.* ACM SIGSOFT Software Engineering Notes, 21:9-12.

Maedche, A. and Staab, S. (2001). *Ontology learning for the semantic web*. Intel. Systems, IEEE, 16:2:72-79.

Nonaka, I. (1994). *A dynamic theory of organizational knowledge creation.* Organization Science, 5(1):14-37.

Novacek, V., Laera, L., and S.Handschuh (2007). *Semiautomatic integration of learned ontologies into a collaborative framework*. In Proceedings of IWOD/ESWC, and ESWC 2007.

Noy, N. and Klein, M. (2004). *Ontology evolution: Not the same as schema evolution*. Knowledge and Information Systems, 6:428–440.

Noy, N. and Musen, M. (2000). *Prompt: Algorithm and tool for automated ontology merging and alignment.* Proceeding of National Conference On Artificial Intelligence.

Nyulas, C., O'Connor, M., and Tu, S. (2007*). Datamaster - a plug-in for importing schemas and data from relational databases into protege*. 10th Intl. Protege Conference, Budapest (2007).

Pressman, R. (2006). *Software Engineering: A Practioners Approach.* Sixth Edition McGraw-Hill, New York 2006.

Ramos, L. and Gil, R. (2007). *Propuesta de sistema de información para apoyar la gestión de la educación a distancia*. CISCI'07, July. Orlando-USA.

Shamsfard, N. and Abdollahzadeh, A. (2003). *The state of the art in ontology learning: a framework for comparison*. Know. Enginer. Rev, 18 (4):293-316.

Simperl, E., Tempich, C., and Vrandecic, D. (2008). *A Methodology for Ontology Learning.* Chapter of Book

Ontology Learning & Population. IOS Press, Buitelaar & Cimiano Eds.
Sommerville, I. (2006). *Software Engineering*. Pearson Education.
Yao, Y., Zeng, Y., Zhong, N., and Huang, X. (2007). *Knowledge retrieval*. In Proceedings IEEE/WIC/ACM International Conference on Web Intelligence.
Zhou, L. (2007). *Ontology learning: state of the art and open issues.* Information Technology and Management, 8(3):241–252.