Integrating Information Extraction, Ontology Learning and Semantic Browsing into Organizational Knowledge Processes

José Iria[1], Fabio Ciravegna[1], Philipp Cimiano[2], Alberto Lavelli[3], Enrico Motta[4], Luca Gilardoni[5], Eddie Moench[6]

[1] Department of Computer Science, University of Sheffield, UK
{j.iria, f.ciravegna}@dcs.shef.ac.uk
[2] AIFB, University of Karlsruhe, Germany
cimiano@aifb.uni-karlsruhe.de
[3] ITC-Irst, Italy
lavelli@itc.it
[4] KMI, Open University, UK
e.motta@open.ac.uk
[5] Quinary, Italy
gil@quinary.com
[6] Ontoprise, Germany
Moench@ontoprise.de

Abstract. Ontology-based approaches to Knowledge Management promise better access to relevant knowledge by providing a domain-specific vocabulary that is used for describing the contents of knowledge as well as for retrieving that knowledge. Despite the potential benefits, ontology-based approaches require a considerable amount of commitment and expertise in tasks like creating and maintaining ontologies, annotating documents and integrating information. We are studying and implementing a methodology for automating such tasks, which makes use of Information Extraction and Integration, Ontology Learning and Semantic Browsing to effectively acquire, share and reuse knowledge in an organizational setting. Applications of the proposed methodology are under development for knowledge management in the legal domain and in the field of biotechnology.

1 Introduction

Managing information is of paramount importance nowadays, especially when it comes to the diffusion and reuse of proprietary knowledge. Employees of large companies and organizations have at their disposal dozens of knowledge repositories located in different parts of the world, accessible through intranets. Nevertheless, knowledge remains hard to capture, share and reuse amongst employees. Capturing knowledge involves the formalization of some work activity, which is itself a non-trivial task. Locating shared information is hard, due to the sheer volume of information available and due to the fact that the access is usually based on keyword searching and browsing. The reuse of knowledge typically involves integrating pieces of information, which is also difficult and time consuming because such pieces of information are likely to be redundant, partial, and stored in different formats. At a time when companies are more and more valued for their "intangible assets" (i.e. their knowledge), unmanageable knowledge may result in the loss of a company's value.

Ontology-based approaches to Knowledge Management [29] promise better access to relevant knowledge by providing a domain-specific vocabulary that is used for describing the contents of knowledge as well as for retrieving that knowledge. Ontologies formalize and structure (part of) the domain of the work activity and thus contribute to streamline the capture of knowledge. Documents (and other work artifacts) are semantically annotated and accessed using the vocabulary provided by the ontology. Whereas traditional keyword search techniques tend to retrieve too much irrelevant information as a term can have different meanings in distinct contexts, searches based on ontological concepts tend to be more effective as they avoid such ambiguity. Despite the potential benefits, ontology-based approaches to Knowledge Management require a considerable amount of commitment and expertise in tasks like creating and maintaining ontologies, annotating documents and integrating information. Methodologies which seek to automate these tasks as much as possible are object of active research. As a matter of fact, some tasks become unmanageable if not supported by automated procedures. For example, manually annotating documents and e-mail produced daily is unthinkable. Other tasks may benefit from automation in different ways. Concretely, automatic procedures can lessen the need for trained human annotators or help in bootstrapping the task, thus reducing the cost of creating knowledge structures.

The Dot.Kom IST project [10] funded within Framework 5 (IST-2001-34038) is studying and implementing a methodology for automating typical tasks of ontology-based approaches to Knowledge Management. We are focusing on the study of the knowledge life-cycle and the way in which Semantic Web Technologies, Natural Language Processing, Information Retrieval, Extraction and Integration, Ontology Learning and Semantic Browsing can be applied to effectively acquire, share and reuse knowledge in an organizational setting. The overall approach builds on ENRICH [22], a methodology that intends to promote organizational learning within an enterprise by providing support for a number of learning processes. The consortium has developed methods for automatic document annotation that are used to populate knowledge bases and for highlighting relevant facts in documents. In addition, Ontology Learning methods were developed in order to address the difficulty in modeling the knowledge of the domain in question. Semantic Browsing builds on traditional web browsing by associating semantics to parts of a document and providing contextualized services that are able to interact with the knowledge bases. Applications of the proposed methodology are under development for knowledge management in the legal domain and in the field of biotechnologies.

In this paper, we present the methodology for Knowledge Management developed within the Dot.Kom consortium, and we briefly describe the software tools developed to support our methodology. The paper is organized as follows. Section 2 reviews the ENRICH methodology, which constitutes the basis for our approach. In Section 3 we discuss the shortcomings of the original methodology and we give an overview of the improvements introduced by Dot.Kom. The following sections detail the several methods and tools that we developed to become part of our methodology. Concretely, Section 4 describes Automatic Semantic Annotation methods, Section 5 details Ontology Learning methods and Section 6 describes Semantic Browsing. Finally, Section 7 presents the conclusions.

2 The ENRICH methodology

The aim of the ENRICH methodology is to promote organizational learning within an enterprise [22]. Learning can be characterised on the level of the individual, group and organisation. For example, a new idea may be initially developed by an individual within a team. This idea may then be modified and elaborated through collaboration within the team, and become incorporated into revised work practices. This may become known and adopted by other teams and eventually even lead to changes in company policy. To reflect that, the ENRICH methodology is based on four main types of organisational learning: reflection-in-action (at the individual level), domain construction and community of practice learning (at the group level) and perspective taking (at the organisational level).

The methodology is centered around Work Representations (WRs). These are documents or other artifacts, which are embedded in work activities that require communication. Working Representations can be divided in subsections and each of them can be linked to a discussion area in which people discuss about potential changes to the actual WR. Hence, the starting point of the ENRICH methodology is a scenario in which members of an organization collaborate and share knowledge by communicating and sharing work representations. To support these activities and to maximize their learning value, ENRICH proposes the notion of Enriched Work Representations (EWRs). These are work representations associated with a representation of the context in which they were created. This context can be informal, such as a record of the discussions which led to the generation of the work representation in question, or formal, such as a formal specification of a design associated with a work representation of the artifact in question. In Dot.Kom, we mainly focus in the use of ontologies as formal representations of the context of WRs, and in annotation of WRs as the basis for knowledge sharing. A central role in the ENRICH methodology is played by the so called 'team developers', who are in charge of developing mechanisms to capture the association between a WR and its context. A team developer is typically in charge of developing an ontology, which is then used to provide the conceptual basis for annotating WRs semantically with regard to the ontology.

Partial support for the ENRICH methodology is provided by a tool [27] that enables semi-automatic mechanisms for coupling of a WR with the ontology which contextualizes it. In addition, several ontology editors exist to support ontology development and ontology-driven annotation of resources [28].

3 Dot.Kom's Approach

Originally, support for the ENRICH methodology was by and large 'manual'. Imagine a team working on an existing Working Representation containing guidelines for nurses about how to take care of patients in English hospitals. Consider that the goal of the team is to produce the Italian equivalent of such a Working Representation, that is, the guidelines (in Italian) for nurses about how to take care of patients in Italian hospitals. Team members carry on discussions concerning a certain guideline in specific discussion areas. The team developer identifies the terminology and concepts used in the discussion areas, creates and maintains the ontology for the particular domain of nursing, and annotates the discussion threads and the guideline documents accordingly. Team members browse the discussion area, and perform keyword searches on past discussion threads.

In Dot.Kom we have identified opportunities for improving and for automating some of the key tasks of ENRICH. Concerning the tasks of the team developer, manual annotation is replaced by automatic or semi-automatic annotation of the WRs; ontology creation is bootstrapped by automatic generation of an initial ontology; ontology maintenance is simplified by the automatic detection of changes in the terminology used in the discussion spaces. Concerning the tasks of the team member, browsing of WRs now integrates browsing of the ontology (instances and related concepts) and access to contextualized services associated with concepts found in the WRs; related discussions can be clustered for inspection; and logical query answering capabilities are being added to searches. More concretely, the research areas we are focusing in the Dot.Kom project are:

- *User-centred semantic annotation*. The consortium has developed systems that learn to annotate WRs and discussion areas by monitoring the annotations inserted by human annotators. When trained, the systems provide suggestions of annotations to the user, who merely checks their correctness. We developed Information Extraction systems [7][8] (detailed in Section 4) and integrated them into our annotations tools [9] [15][32] and our semantic browsing tool [13]. The tools are designed with normal users in mind, requiring no knowledge about IE or ontology formalisms.
- *Unsupervised annotation of large repositories via Harvesting*. To support the annotation of information that is dispersed throughout multiple potential sources, the consortium has been developing a methodology and a system [11] (detailed in Section 4) that automatically annotates domain-specific information in large repositories, such as the Web or a company's repository, with minimum user intervention and high portability across domains. The system is able to relate and collate information, and create a domain-specific knowledge base of facts.
- *Semi-automatic ontology construction and maintenance*. Ontology learning and text mining [4][5][6] methods developed by the consortium (detailed in Section 5) are employed to support the team developer in building and maintaining ontologies. Applying ontology learning methods to the WRs serves two purposes. First, a 'kickoff' ontology can be automatically generated and manually refined afterwards. Second, team developers can be alerted of changes in the terminology used in WRs, which gives an hint that the ontology no longer adequately represents the context for the WRs.
- *Semantic browsing*. Semantic browsing provides a semantics layer on top of classic browsing of WRs. This means that the visual representation of the WR is enriched by highlighting concepts of interest and, most notably, by making available contextual services related to those concepts to the user. Information Extraction and Semantic Web Services provide the underlying technologies for our Semantic browsing tool [13] (detailed in Section 6). We are currently adding support for automatic semantic annotation directly into the tool, which is expected to replace the annotation tools altogether in the near future.
- *Logical query answering*. Besides semantic browsing, a natural way of taking advantage of the annotations introduced into the WRs is to support logical queries. The consortium has developed a tool [21] that uses an inference engine in order to perform semantic retrieval. We are currently considering using a natural language front end [3][31] as well.

Figure 1 illustrates the research areas together with the respective tools developed by the Dot.Kom consortium and how they improve the original ENRICH methodology. The rest of the paper is devoted to providing details about each one of the research areas mentioned, namely Automatic Semantic Annotation of Documents, Ontology Learning and Semantic Browsing.

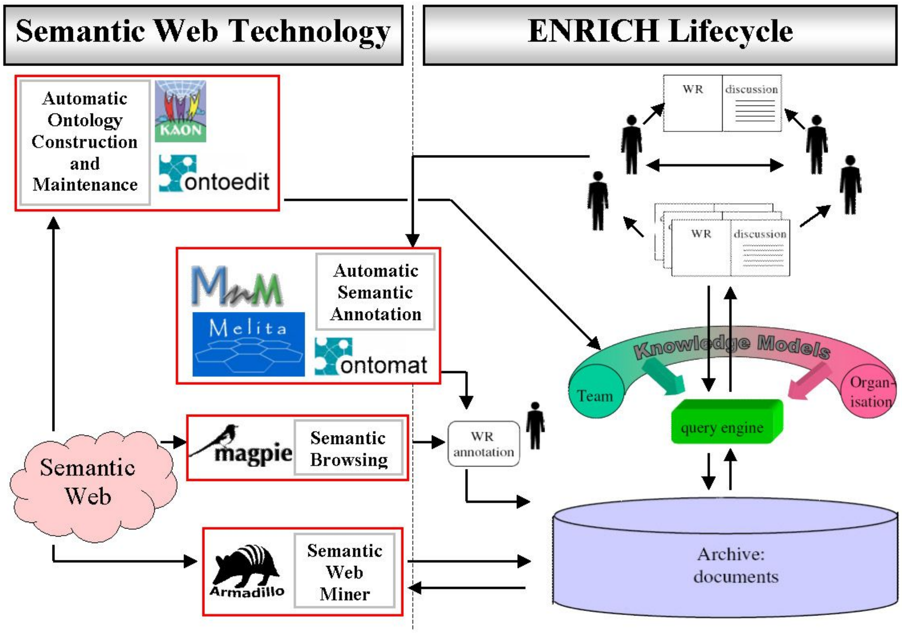

Fig. 1. Conceptual view of Dot.Kom's research areas and their application in the context of the ENRICH methodology

4 Automatic Semantic Annotation

In order to support automatic semantic annotation, we have developed three main approaches: user centred and supervised [7][8][9], unsupervised [11] and pattern-based [2]. The first two are supported by Automatic Information Extraction from Documents with Machine Learning (Adaptive IE in the following). Adaptive IE has been researched in areas such as Information Extraction [23] and Wrapper Induction [19]. In the Dot.Kom project, we focus on adaptive, machine learning-based methods. An important requisite for integrating IE into end-user tools (e.g. the semantic browser) is that the IE methods must be able to adapt to a large variety of application domains. We have designed and implemented two state-of-the-art algorithms [7][8] that handle any domain that requires the identification of entities and implicit relations between those entities. By 'implicit' we mean that entities identified in the text are explicitly stated to play a specific role in a relation but the relations are not explicitly modeled. To exemplify, a typical benchmark task for implicit relation extraction is the identification of 'location', 'speaker', 'start time' and 'end time' of a seminar where only one seminar announcement is present per document (the simplifying assumption) - were the documents to contain more than one announcement per document, current state-of-the-art systems would be unable to determine which 'speaker' relates to which seminar and so forth. Although this is a limitation, the Amilcare system [7] has been integrated in Armadillo, Melita, MnM, OntoMat, SemantiK and OntoAnnotate, released to about 50 industrial and academic sites and used in real world applications (see `http://nlp.shef.ac.uk/amilcare/` for a complete list). Ties [8] has been integrated with SemantiK and OntoAnnotate.
The next step in this direction is modelling explicit relations. Explicit relation extraction is more general, in the sense that relations between entities are actually modeled and learned.
Recent machine learning approaches [33] to relation extraction define kernel methods for support vector machines (SVM) [30]. Our approach identifies three major components of research: corpus representation, pattern mining and classification. We are experimenting with a uniform graph representation that encompasses the linguistic (e.g. lemma, POS) and formatting information (e.g. HTML tags) of the text together with the user annotations linked to the domain ontology. Using this kind of representation makes the algorithm adaptable to several document types and requires no special-case treatment for ontological information. The goal of relation extraction is to mine subgraph patterns that discriminate between the subgraphs around user annotations (positive subgraphs) and the rest of the graph (negative subgraphs), and then use the mined patterns as features for a classification task. We are adapting state-of-the-art frequent subgraph miners [18] to take into account posit-

ive/negative subgraphs. SVM is used for the classification task, as it is well-suited to operate in sparse and high-dimensional datasets.

Information Extraction is rather more concerned with identifying information in selected documents than with searching for and integrating information dispersed throughout innumerous documents. As such, we have further developed a methodology and a system [11] that automatically annotates on large repositories. We use it to harvest the Web and the company's repositories to find instances of concepts of the domain of a WR. Our system finds multiple potential sources for the information and applies IE to extract it, annotates it and integrates it in a repository. The system achieves this in a largely unsupervised way by taking advantage of the redundancy of information in such large repositories, making it quite easily adaptable to new domains. We are currently addressing two key challenges. First, information coming from diverse sources poses the problem of integration, as it is likely to be redundant, partial, have different superficial formats and even be contradictory. Ad-hoc methods are being replaced by more sound information retrieval and learning-based methods for information integration. Second, even though the methodology ensures adaptability to new domains, there are practical issues that still require an expert in the internals of the system. We are currently working on a modular, Semantic Web Services-based architecture that will allow our system to become an off-the-shelf component for other Knowledge Management tools.

The third approach to semantic annotation is a pattern-based one which relies on the Google API to match Hearst-style lexico-syntactic patterns in the WWW indicating an instance-of relationship and annotate instances in a web page with the concept maximizing the number of corresponding patterns found [2]. This approach called **P**attern-based **An**notation through **K**nowledge **o**n the **W**eb (PANKOW) has been implemented in OntoMat [15], one of the consortium's annotation tools. The approach is unsupervised in the sense that it relies only on a fixed set of core patterns, as well as domain-independent as it can be used with any ontology. Currently, we are integrating PANKOW with MagPie, a semantic web browser. In the future, we expect that annotations created with PANKOW can be used to train and bootstrap an adaptive information extraction system such as Amilcare [7].

5 Ontology Learning

Several methods have been proposed in the literature to automate the ontology engineering process. In Dot.Kom we focus particularly in automating the acquisition of taxonomies (or concept hierarchies) from text. For this purpose, we exploit two types of approaches: one based on a vector-space model and clustering techniques and one exploiting different sources of knowledge. Both approaches are implemented in the TextToOnto Ontology Learning Framework [17].

The approaches based on a vector-space model represent a word or a term as a vector containing features derived from the text and make use of clustering techniques to produce clusters of similar terms. Similarity-based methods [24], on the one hand, use similarity/distance measures in order to compute the pairwise similarity/distance between vectors corresponding to two terms in order to decide if they should be clustered or not. Set-theoretical approaches [25], on the other hand, partially order the objects according to the inclusion relations between their feature sets.

Our approach, detailed in [5][6], consists of a novel set-theoretical method that relies on Formal Concept Analysis [14], a method mainly used for the analysis of data, and on the distributional hypothesis, that is, terms are considered to be similar to the extent to which they share contexts. Further, we assume that verbs pose more or less strong selectional restrictions on their arguments. The concept hierarchy is built via Formal Concept Analysis using syntactic verb-argument dependencies, namely verb/PP-complement, verb/object and verb/subject dependencies as formal context. For each noun appearing as head of these argument positions we use the corresponding verbs as features. Evaluation consists in comparing the produced concept hierarchies against handcrafted taxonomies from different domains. The results show that our method performs well compared to other state-of-the-art algorithms [6].

The other approach we use relies on the combination of different approaches and sources of knowledge such as WordNet, patterns matched in the Web as well as in a domain-specific corpus and the distributional similarity of terms [4]. We are currently exploring different strategies and machine learning techniques to combine the information from these different approaches and sources of knowledge to decide whether two terms stand in an is-a relation or not with the aim of producing a taxonomy between a given set of terms relevant for the domain or application in question.

In Dot.Kom, both approaches allow team developers to quickly bootstrap the knowledge lifecycle by automatically deriving a “kickoff” ontology from the WRs. We have further implemented methods to assess the similarity between two given ontologies, namely using the Taxonomic Overlap (TO) and Relational Overlap (RO) measures as presented in [20]. We use these methods to evaluate how well the automatically learned ontology conceptualizes the domain. When the similarity measure between a periodically derived ontology and the actual ontology drops bellow a given threshold, the team developer is alerted that the ontology should probably be modified. Again, in modifying the ontology the team developer may take advantage of the automatically derived ontology.

6 Semantic Browsing

Browsing work representations is perhaps the most common activity in the Knowledge life-cycle. The Dot.Kom project has researched on innovative methods that intend to improve the productivity of knowledge workers while browsing. Our approach quite naturally takes advantage of the infrastructure described in previous sections. Automatic semantic annotation, on the one hand, virtually introduces a semantic layer (of metadata) onto WRs, which can be used to facilitate the interpretation of their content in a number of ways. The use of ontologies as a formal representation of the context of WRs, on the other hand, makes it possible to define knowledge services that are able to relate (via relations defined in the ontology) the content in the WR with knowledge stored in the knowledge bases.

The consortium has developed a tool [13] that makes use of the automatic semantic annotation methods on-the-fly in order to associate meaning with pieces of information found in WRs and then, on the basis of the identified meaning, to offer or invoke relevant contextual services and appropriate functionalities. It is important to stress that the tool is not bound to use a predefined ontology or set of services, making it adaptable to a large number of applications. The tool facilitates the interpretation of the content of the WR by highlighting occurrences of concepts of interest. In addition, each such concept may provide an applicable set of services, dynamically evaluated by querying an ontology server, that are made available to the user via a context menu. By providing contextual services, the tool acts as a complementary knowledge source, which team members can call upon to gain instantaneous access to the background knowledge relevant to an identified entity.

Our approach to semantic browsing supports what Boland and Tenkasi [1] term as “perspective taking”. Perspective taking describes the process by which teams in the organization recognise, use and evaluate the perspectives of other teams as part of their work, and use these to reflect on their own work practices. For example, the personal home page of the member of an institution is normally created within the context of that person’s affiliation and organizational role. Some readers might be very familiar with such context, while others might not. In the latter case, the use of our semantic browsing tool is especially beneficial, given that the context is made explicit via highlighting of concepts and the contextual services provided. Because different readers show differing familiarity with the information in a WR and with the relevant background domain, they require different level of sensemaking support. Recently, the tool has been extended [12] to provide support to collaborative browsing activities, as team members involved in a joint information gathering exercise will wish to share knowledge about the web pages visited and the contents found. Additionally, we are looking into applying clustering algorithms to the discussion areas in order to obtain clusters of related discussion areas. The user may then inspect those clusters deemed relevant as part of the browsing activity.

7 Conclusions

In this paper we have presented the methodology for ontology-based Knowledge Management developed in the EU Dot.Kom project. The proposed methodology improves an existing successful organizational learning framework by automating the tasks of creating and maintaining ontologies using Ontology Learning methods and of annotating documents using Information Extraction, as well as by providing advanced support to knowledge retrieval with Semantic Browsing. The support covers most of the knowledge lifecycle.

Evaluation of the methodology comprises two phases. The first phase, recently completed, consists in the evaluation of the individual technologies separately. The second phase, currently underway, consists in applying the methodology as a whole to real-world scenarios via the consortium's industrial partners. For that purpose, applications of the methodology are under development for knowledge management in the legal domain and in the field of biotechnologies.

References

1. R. Boland, R. Tenkasi. Perspective making and perspective taking in communities of knowing. Organization Science, 6 (4) 350-372, 1995.
2. P. Cimiano, S. Handschuh, S. Staab. Towards the Self-Annotating Web, In Proceedings of the 13th World Wide Web Conference (WWW'04), New York 2004.
3. P. Cimiano, ORAKEL: A Natural Language Interface to an F-Logic Knowledge Base, In Proceedings of the 9th International Conference on Natural Language for Information Systems (NLDB'04), Manchester 2004.
4. P. Cimiano, A. Pivk, L. Schmidt-Thieme, S. Staab, Learning Taxonomic Relations from Heterogeneous Evidence, In Proceedings of the ECAI'04 Workshop on Ontology Learning and Population, Valencia 2004
5. P. Cimiano, A. Hotho, S. Staab. Clustering Ontologies from Text. In Proceedings of the Conference on Lexical Resources and Evaluation (LREC), Lissabon 2004.
6. P. Cimiano, A. Hotho, S. Staab, Comparing Conceptual, Partitional and Agglomerative Clustering for Learning Taxonomies from Text, In Proceedings of the European Conference on Artificial Intelligence (ECAI'04). 2004.
7. F. Ciravegna and Y. Wilks. Designing Adaptive Information Extraction for the Semantic Web in Amilcare, in S. Handschuh and S. Staab (eds), Annotation for the Semantic Web, in the Series Frontiers in Artificial Intelligence and Applications by IOS Press, Amsterdam, 2003.
8. http://tcc.itc.it/research/textec/tools-resources/ties.html
9. F. Ciravegna, A. Dingli, D. Petrelli and Y. Wilks. User-System Cooperation in Document Annotation based on Information Extraction. Asuncion Gomez-Perez, V. Richard Benjamins (eds.): Knowledge Engineering and Knowledge Management (Ontologies and the Semantic Web), Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02), 1-4 October 2002 - Sigüenza (Spain), Lecture Notes in Artificial Intelligence 2473, Springer Verlag.
10. http://www.dot-kom.org
11. F. Ciravegna, S. Chapman, A. Dingli, Y. Wilks. Learning to Harvest Information for the Semantic Web. Proceedings of the 1st European Semantic Web Symposium, Heraklion, Greece, May 10-12, 2004.
12. J. Domingue, M. Dzbor. Collaborative Semantic Web Browsing with Magpie. In Proceedings of the 1st European Semantic Web Symposium (ESWS) , Greece, 2004.
13. M. Dzbor, J. Domingue, E. Motta. Towards a Semantic Web Browser. In Proceedings of the 2nd International Semantic Web Conference, Lecture Notes in Computer Science, 2870:690-705, Springer Verlag, 2003.
14. B. Ganter and R. Wille. Formal Concept Analysis - Mathematical Foundations. Springer Verlag, 1999.
15. S. Handschuh, S. Staab, A. Maedche. CREAM - Creating relational metadata with a component-based, ontology-driven annotation framework. ACM K-CAP 2001, Vancouver.
16. S. Handschuh, S. Staab, F. Ciravegna. S-CREAM - Semi-automatic CREAtion of Metadata. In Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02), 2002.
17. The TextToOnto Ontology Learning Framework: http://sourceforge.net/projects/texttoonto/
18. M. Kuramochi and G. Karypis. Finding Frequent Patterns in a Large Sparse Graph, SIAM International Conference on Data Mining, 2004.
19. N. Kushmerick, D. Weld, and R. Doorenbos. Wrapper induction for information extraction. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 1997.
20. A. Maedche, S. Staab. Measuring Similarity between Ontologies. In Proceedings of the European Conference on Knowledge Acquisition and Management, 2000.
21. E. Moench, M. Ullrich, H.-P. Schnurr, and J. Angele. SemanticMiner - Ontology-Based Knowledge Retrieval. In Journal of Universal Computer Science, 9(7):682–696, July 2003.
22. P. Mulholland, Z. Zdrahal, J. Domingue, M. Hatala. A methodological approach to supporting organizational learning. In International Journal of Human-Computer Studies 55, 2001.
23. M. T. Pazienza, editor. Information Extraction: A multidisciplinary approach to an emerging information technology. Springer Verlag, 1999.
24. F. Pereira, N. Tishby, and L. Lee. Distributional clustering of English words. In Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics, pages 183-190, 1993.
25. W. Petersen. A set-theoretical approach for the induction of inheritance hierarchies. Electronic Notes in Theoretical Computer Science, 51, 2002.
26. S. Staab, R. Studer, Y. Sure. Knowledge Processes and Meta Processes in Ontology-based Knowledge Management. In Handbook of Knowledge Management, Springer Verlag, 2002.
27. T. Sumner, and S. Buckingham Shum. From Documents to Discourse: Shifting Conceptions of Scholarly Publishing. Proc. CHI 98: Human Factors in Computing Systems, (Los Angeles, CA), 95-102. ACM Press: NY. 1998.
28. Y. Sure et al. OntoEdit: Collaborative ontology development for the Semantic Web. In Proceedings of the ISWC 2002, June 9-12 2002, Sardinia, Italia., 2002.
29. Y. Sure, S. Staab, R. Studer. Methodology for Development and Employment of Ontology-based Knowledge Management Applications. In S. Staab & R. Studer (eds.) Handbook on Ontologies. Springer 2004.
30. V. Vapnik. Statistical Learning Theory. Wiley-Interscience, New York, (1998).
31. M. Vargas-Vera, E. Motta, J. Domingue. AQUA: An Ontology-Driven Question Answering System. AAAI Symposium on New Directions of Question Answering, Stanford University, 24-26 March, 2003.
32. M. Vargas-Vera, E. Motta, J. Domingue. M. Lanzoni, A. Stutt, F. Ciravegna: MnM: Ontology Driven Semi-Automatic and Automatic Support for Semantic Markup, In Proceedings of the 13th International Conference on Knowledge Engineering and Management (EKAQ 2002), Springer Verlag, 2002
33. D. Zelenko, C. Aone, and A. Richardella. Kernel methods for relation extraction. Journal of Machine Learning Research, 2003.