On-To-Knowledge in a Nutshell

Dieter Fensel, Frank van Harmelen, Ying Ding, Michel Klein, Hans Akkermans
Free University Amsterdam VUA, Division of Mathematics and Informatics
De Boelelaan 1081a, NL-1081 HV Amsterdam, The Netherlands
http://www.ontoknowledge.org; contact: dieter@cs.vu.nl
Jeen Broekstra, Arjohn Kampman, Jos van der Meer, *AIdministrator, The Netherlands*
York Sure, Rudi Studer, *AIFB, University of Karlsruhe, Germany*
Uwe Krohn, John Davies, *BT, Ipswich, UK*
Robert Engels, *CognIT, Oslo, Norway*
Victor Iosif, *Enersearch AB, Gothenburg, Sweden*
Atanas Kiryakov, *OntoText, Sofia, Bulgaria*
Thorsten Lau, Ulrich Reimer, *Swiss Life, Zurich, Switzerland*
Ian Horrocks, *University of Manchester, UK*

1. Introduction

The World Wide Web (WWW) has drastically changed the availability of electronically available information. Currently, there are around 3 billion documents in the WWW, which are used by more than 500 million users internationally. And that number is growing fast. However, this success and exponential growth makes it increasingly difficult to find, to access, to present, and to maintain the information required by a wide variety of users. Ontologies and semantic web technology will improve this situation [Fensel, 2001]. *On-To-Knowledge* builds an ontology-based tool environment to speed up knowledge management, dealing with large numbers of heterogeneous, distributed, and semi-structured documents typically found in large company intranets and the World Wide Web. The project's target results are: (1) a toolset for semantic information processing and user access; (2) OIL, an ontology-based inference layer on top of the World Wide Web; (3) an associated methodology and validation by industrial case studies. This paper offers an overview of the *On-To-Knowledge* approach to knowledge management.

2. Tool Environment for Ontology-based Knowledge Management

A key deliverable of the On-To-Knowledge project is the resulting software toolset. Several consortium partners are participating in the effort to realize in software the underpinning ideas and theoretical foundations of the project. A major objective of the project is to create intelligent software to support users in both accessing information and in the maintenance, conversion, and acquisition of information sources. These tools are based on a three-layered architecture around information access, information storage, and information generation. Most of the tools presented here in Figure 1 are described below.

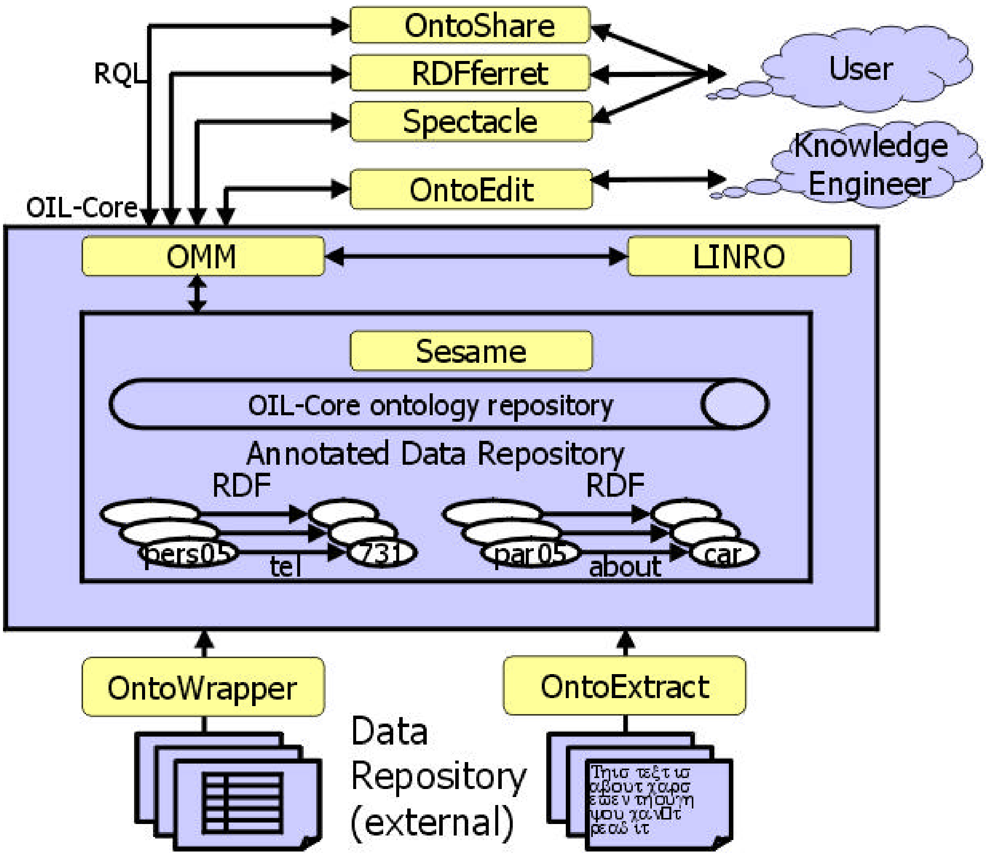

Figure 1. The technical architecture of On-To-Knowledge.

Observing that RDF-annotated information resources are likely to be complemented by non-annotated information for a considerable period to come, and that any given RDF description of a set of resources will give one particular perspective on the information described, **RDF*ferret*** combines full text searching with RDF querying. **OntoShare** enables the storage of best practice information in an ontology and the automatic dissemination of new best practice information to relevant co-workers. It also allows users to browse or search the ontology in order to find the most relevant information to the problem that they are dealing with at any given time. The ontology helps to orientate new users and acts as a store for key learning and best practices accumulated through experience. **Spectacle** organizes the presentation of information. This presentation is ontology driven. Ontological information, such as classes or specific attributes of information, is used to generate exploration contexts for users. An exploration context makes it easier for users to explore a domain. The context is related to certain tasks, such as finding information or buying products. The context consists of three modules:

- Content: specific content needed to perform a task,
- Navigation: suitable navigation disclosing the information,
- Design: applicable design displaying the selected content.

OntoEdit makes it possible to inspect, browse, codify and modify ontologies, and thus serves to support the ontology development and maintenance task. Modelling ontologies using OntoEdit involves modelling at a conceptual level, viz. (i) as independently of a concrete representation language as possible, and (ii) using GUI's representing views on conceptual structures (concepts, concept hierarchy, relations, axioms) rather than codifying conceptual structures in ASCII. The **Ontology Middleware Module** (OMM) can be seen as the key integration component in the OTK technical solution architecture. It supports well-defined application programming interfaces (OMAPI) used for access to knowledge and deals with such matters as:

- Ontology versioning, including branching.

- Security – user profiles and groups are used to control the rights for access, modifications, and publishing;
- Meta-information and ontology lookup – support for meta-properties (such as Status, Last-Updated-By, Responsible, Comments, etc.) for whole ontologies, as well as for separate concepts and properties. Ontology and concept lookup according to meta-information is possible.
- Access via a number of protocols: HTTP, RMI, EJB, CORBA, and SOAP.

Sesame is a system that allows persistent storage of RDF data and schema information and subsequent online querying of that information. Sesame has been designed with scalability, portability and extensibility in mind. One of the most prominent modules of Sesame is its query engine. This query engine supports an OQL-style query language called RQL. RQL supports querying of both RDF data (e.g. instances) and schema information (e.g. class hierarchies, domains and ranges of properties). RQL also supports path-expressions through RDF graphs, and can combine data and schema information in one query. **LINRO** provides additional reasoning services so as to extend the functionality provided by SESAME. Most of the classical reasoning tasks for description logics (DL) are available, including realization and retrieval. The goal was to enable even wider set of applications, such as information extraction and automatic ontology integration. A strategy called pre-reasoning was used to implement workarounds for a number of logical problems proven to be computationally intractable for languages as expressive as OIL.

Information extraction and ontology generation is performed by means of the **CORPORUM** toolset and is situated in the *extraction layer*. CORPORUM has two related, though different, tasks: interpretation of *natural language* texts and extraction of *specific information* from free text. Whereas CORPORUM tools can perform the former process autonomously, the latter task requires a user who defines business rules for extracting information from tables, (phone) directories, home-pages, etc. Although this task is not without its challenges, most effort focuses on the former task, which involves natural language interpretation on a syntactic and lexical level, as well as interpretation of the results of that level (discourse analysis, co-reference and collocation analysis, etc.).

3. OIL: Inference Layer for the Semantic World Wide Web

The OIL language (cf. [Fensel et al., 2001]) is to designed to combine frame-like modeling primitives with the increased (in some respects) expressive power, formal rigor and automated reasoning services of an expressive description logic. OIL also comes "web enabled" by having both XML and RDFS based serializations (as well as a formally specified "human readable" form, which we will use here[1]). As part of the Semantic Web activity of the W3C, a very simple web-based ontology language had already been defined, namely RDF Schema. This language only provides facilities to define class- and property-names, inclusion axioms for both classes and properties (subclasses and subproperties), and to define domain and range constraints on properties. OIL has been designed to be a superset of the constructions in RDF Schema and the syntax of OIL has been designed such that any valid OIL document is also a valid RDF(S) document when all the elements from the OIL-namespace are ignored. The RDF Schema interpretation of the resulting subdocument is guaranteed to be sound (but of course incomplete) with respect to the interpretation of the full OIL document.

For many of the applications, it is unlikely that a single language will be ideally suited for all uses and all users. In order to allow users to choose the expressive power appropriate to their application, and to allow for future extensions, a layered family of OIL languages has been described. The sub-language OIL Core has been defined to be exactly the part of OIL that coincides with RDF(S). This amounts to full RDF(S), without some of RDF's more dubious constructions: containers and reification.

[1] http://www.ontoknowledge.org/oil/syntax/

Meanwhile, OIL has been adopted by a joined EU/US initiative that developed a language called DAML+OIL[2]. In November 2001, the W3C started a Web Ontology Working Group for defining a language. This group is chartered to take DAML+OIL as its starting point.

4. Business Applications in Semantic Information Access

Tool development in On-To-Knowledge was strongly driven by several case studies. Some of them started in the earliest stages of the project so that the project was geared towards practical demands from the very beginning. The case studies also serve as a means to evaluate the project results . Swiss Life carried two of the case studies out. One of these approached the problem of finding relevant passages in a very large document about the International Accounting Standard (IAS) on the extranet (over 1000 pages). Accountants who need to know certain aspects of the IAS accounting rules use this document. As the IAS standard uses very strict terminology, it is only possible to find relevant text passages when the correct terms are used in the query. Very often, this leads to poor search results. Although the ontology that was built is structurally quite simple, it greatly improves search results. Swiss Life's second case study is a skills management application that uses manually constructed ontologies about skills, job functions, and education. These consist of 800 concepts with several attributes, arranged into a hierarchy of specializations. There are also semantic associations between these concepts. The skills management system makes it easy for employees to create in a personal home page on the company's intranet that includes information about personal skills, job functions, and education. The case study done by EnerSearch AB focuses on validating the industrial value of the project's results with respect to the needs of a virtual organization. The goal of the case study is to improve knowledge transfer between EnerSearch's in-house researchers and outside specialists via the existing website. The study also aims to help the partners from shareholding companies to obtain up-to-date information about research and development results.

Finally, a methodology for employing ontology-based tools for knowledge management applications was developed by applying an initial baseline methodology (cf. [Staab et al., 2001]) in the case studies and continuously improving it with the insight gained from experience. This methodology provides guidelines for introducing knowledge management concepts and tools into enterprises, helping knowledge providers and seekers to present knowledge efficiently and effectively. In On-To-Knowledge, ontologies are the core element. Thus, the methodology has a strong focus on ontology development.

5. Conclusions

The Web and company intranets have boosted the potential for electronic knowledge acquisition and sharing. Given the sheer size of these information resources, there is a strategic need to move up in the data – information – knowledge chain. *On-To-Knowledge* takes a necessary step in this process by providing innovative tools for semantic information processing and thus for much more selective, faster, and meaningful user access. The technology developed has been proven to be useful in a number of case studies. We can improve information access in the large intranets of sizeable organizations. The technology has been used to facilitate electronic knowledge sharing and re-use for customer relationship management and knowledge management in virtual organizations.

More information on On-to-knowledge can be found at [Davies et al., to appear] and on the project's website www.ontoknowledge.org.

[2] http://www.daml.org.

References

[David et al., to appear] J. Davies et al. (eds.): *The Semantic Web: Ontology-driven Knowledge Management*, John Wiley, to appear.

[Fensel, 2001] D. Fensel: *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*. Springer-Verlag, Berlin, 2001.

[Fensel et al., to 2001] D. Fensel, I. Horrocks, F. van Harmelen, D. McGuinness, and P. F. Patel-Schneider: OIL: Ontology Infrastructure to Enable the Semantic Web, IEEE Intelligent System, 16(2), 2001.

[Staab et al., 2001] S. Staab, H.-P. Schnurr, R. Studer, and Y. Sure: Knowledge Processes and Ontologies, *IEEE Intelligent Systems*, 16(1), 2001.