On2broker in a Nutshell

Dieter Fensel, Andreas Witt, Jürgen Angele, Stefan Decker, Michael Erdmann, Hans-Peter Schnurr, Steffen Staab, and Rudi Studer

Institute AIFB, University of Karlsruhe, D-76128 Karlsruhe, Germany
dfe@aifb.uni-karlsruhe.de, http://www.aifb.uni-karlsruhe.de/www-broker

Abstract. On2broker provides brokering services to improve access to heterogeneous, distributed and semistructured information sources as they are presented in the World Wide Web. It relies on the use of *ontologies* to make explicit the semantics of web pages, to formulate queries and to derive answers for them. In the poster we will sketch its main purpose and the general architecture of On2broker.

In the area of knowledge-based systems *ontologies* have been developed for structuring and reusing large bodies of knowledge (cf. CYC [Lenat, 1995], KIF/Ontolingua [KIF], $(KA)^2$ [Benjamins et al., 1998]). Ontologies are consensual and formal specification of a vocabulary used to describe a specific domain. On2broker uses such ontologies to improve *access to information* provided in intranets and in the internet. It uses semantic information for guiding the query answering process. It provides the answers with a well-defined syntax and semantics that can be directly understood and further processed by automatic agents or other software tools. It enables a homogeneous access to information that is physically distributed and heterogeneously represented in the WWW and it provides information that is not directly represented as facts in the WWW but which can be derived from other facts and some background knowledge. This tool service can also be used to create and maintain such semistructured information sources, i.e. it is a tool for web site construction and restruction. *Automatic document generation* extracts information from weakly structured text sources and creates new textual sources. The gist of this application is that it generates semistructured information presentations *from* other semistructured ones. *Maintenance* of weakly structured text sources helps to detect inconsistencies among documents and to detect inconsistencies between documents and external sources, i.e., to detect incorrectness. The overall picture of On2broker is provided in Figure 1 which includes four basic engines representing different aspects.

- The **info agent** is responsible for collecting factual knowledge from the web using various style of meta annotations, direct annotations like XML and in future also text mining techniques.
- The **inference agent** uses facts and ontologies to derive additional factual knowledge that is only provided implicitly. It frees knowledge providers from the burden of specifying each fact explicitly.
- The **query engine** receives queries and answers them by checking the content of the databases that were filled by the info and inference agents.

- The **database manager** is the backbone of the entire system. It receives facts from the Info agent, exchanges facts as input and output with the inference agent, and provides facts to the query engine.

The strength of Ontobroker is the tight coupling of informal, semiformal and formal information and knowledge. This supports their maintenance and provides a service that can be used more generally for the purpose of *knowledge management* and for integrating knowledge-based reasoning and the semiformal representation of documents.

On2broker is available on the web and has been applied in a number of applications in the meantime. The most prominent one is the $(KA)^2$ initiative that develops an ontology for annotating web documents of the knowledge acquisition community [Benjamins et al., 1998].

References

[Benjamins et al., 1998] R. Benjamins, D. Fensel, and A. Gomez Perez: Knowledge Management Through Ontologies. In *Proceedings of the Second International Conference on Practical Aspects of Knowledge Management (PAKM'98)*, Basel, Schweiz, 29.-30. Oktober 1998.

[KIF] Knowledge Interchange Format. Draft proposed American National Standard (dpANS) NCITS.T2/98-004. http://logic.stanford.edu/kif/dpans.html.

[Lenat, 1995] D. B. Lenat: CYC: A Large-Scale Investment in Knowledge Infrastructure, *Communications of the ACM* 38(11), 1995.

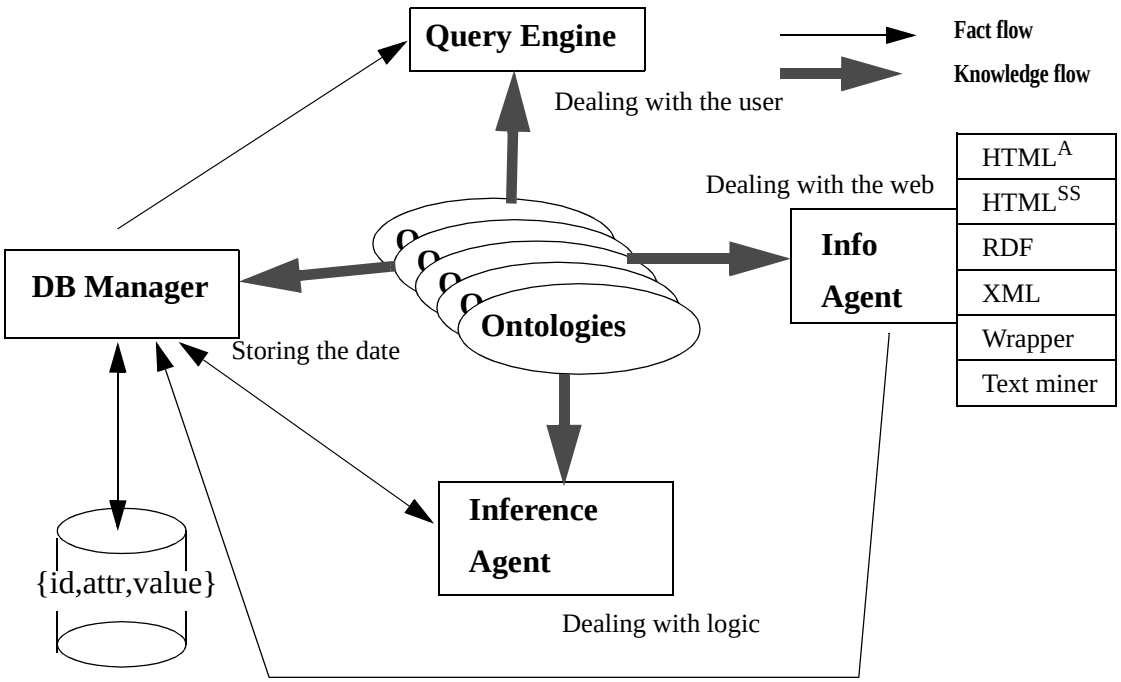

Figure 1 On2brokers Architecture.