An Information Food Chain for Advanced Applications on the WWW

Stefan Decker[1], Jan Jannink[1], Sergey Melnik[1], Prasenjit Mitra[1], Steffen Staab[2], Rudi Studer[2], and Gio Wiederhold[1]

[1] Computer Science, Stanford University, Stanford, CA, 94305, U.S.A.
[2] University of Karlsruhe, Institut AIFB, 76128 Karlsruhe, Germany
{stefan, jan, melnik, prasen9, gio}@db.stanford.edu
{staab, studer}@aifb.uni-karlsruhe.de

Abstract. The growth of the WWW has resulted in amounts of information beyond what is suitable for human consumption. Automated information processing agents are needed. However, with the current technology it is difficult and expensive to build automated agents. To facilitate automated agents on the web we present an *information food chain* for advanced applications on the WWW. Every part of the food chain provides information that enables the existence of the next part.

1 Introduction

The growth of the World Wide Web has resulted in large amounts of information available for human consumption. Since humans have a limited capacity for processing information, we need automated information processing agents [GK94]. However, with the current technology it is difficult and expensive to build automated agents, because agents are not able to understand the meaning of the natural language terms found on today's webpages. To facilitate automated agents on the web, agent interpretable data is required. Creating and deploying data about a particular domain is a high effort task, and it is not immediately clear how to support this task. This paper presents an *information food chain* [E97] for advanced applications on the WWW. Every part of the food chain provides information that enables the existence of the next part.

2 The Information Food Chain

For data exchange on the web, it is necessary to have a specification of the terminology of the domain of interest. *Ontologies* [FH97] are a means for knowledge sharing and reuse, and capture the semantics of a domain of interest. An ontology is a formal specification of vocabularies used to describe a specific domain. It provides a basis for a community of interest for information exchange. Since there will be no ontology available describing all possible domains, multiple ontologies for different application domains are needed. Ontologies have been defined using

RDF [LS99], RDF Schema [BG99], XOL, [KCT99]. An XML-based representation language and formal ontologies are a foundation for automated agents on the web. However, to create machine-interpretable data and deploy it, we need infrastructure (e.g. support- and deployment tools), given by an information food chain (see Fig. 1).

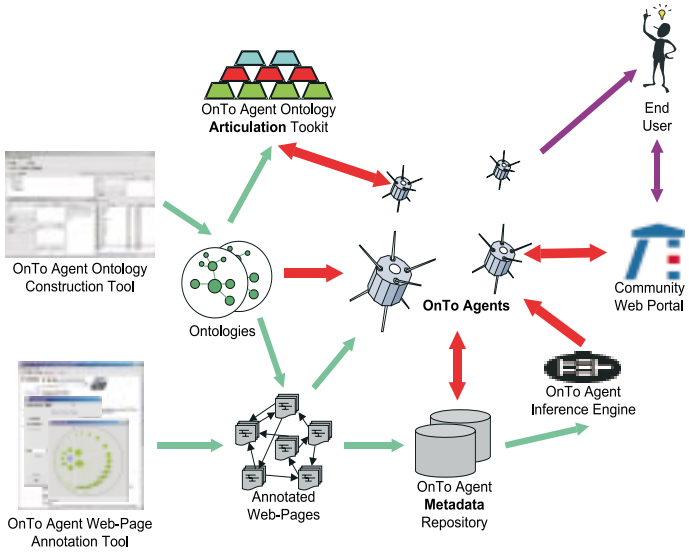

Fig. 1. Agents Information Food Chain

Ontology Construction Tool The food chain starts with the construction of an ontology. Constructing and maintaining an ontology involves human interaction. Also ontologies evolve and change over time (as our knowledge, needs, and capabilities change) so reducing acquisition and maintenance cost of ontologies is an important task. An *Ontology Construction Tool* is necessary to provide the means to construct ontologies in a cost-effective manner. Examples of ontology editors are e.g. the Protégé framework for knowledge-base system construction [1], and the WebOnto-Framework [D98]. WebOnto supports collaborative, distributed browsing, creation and editing of ontologies by providing a direct manipulation interface that displays ontological expressions. However, none of those tool is yet used for creating ontologies for web based agent applications.

Webpage Annotation Tool For information on webpages to be machine-interpretable, semantic information about the content of the page is needed. A *Webpage Annotation Tool* provides an annotator the means to browse an ontology and to select appropriate terms of the ontology and map them to sections of a webpage.

[1] http://smi-web.stanford.edu/projects/protege/

Using Ontopad [DEF99], an enhanced HTML editor, the annotator can select a portion of the text from a webpage and choose to add a semantic annotation, which is inserted into the HTML source. However, for significant annotation tasks a practical tool also has to exploit information extraction techniques for semi-automatic metadata creation. Often it is sufficient to give the user a choice to annotate a phrase with some ontological expressions. Resources like WordNet [F98] and results obtained from the Scalable Knowledge Composition (SKC) project [MGK00] can be used to annotate webpages semi-automatically.

Ontology-Articulation Toolkit In order to solve a task that involves consulting multiple information sources that have unknown ontologies, an automated agent needs to bridge the semantic gap between the known and the unknown ontologies. In the Scalable Knowledge Composition Project (SKC [MGK00]) we have developed tools that automatically generate articulations or semantic bridges, among multiple ontologies (by consulting online dictionaries and using other heuristics), and presents it to an expert for validation.

Agents Inference System For declarative information processing an agent needs an *Inference Engine* for the evaluation of rules and queries. An inference engine helps to exploit available metadata by infering further implicit metadata [DBS98]. The properties of the reasoning capabilities have to be carefully chosen. A reasoning mechanism, which is too powerful, has intractable computational properties, whereas a too limited approach does not enable the full potential of inference. Deductive database techniques have proven to be a good compromise between these tradeoffs.

Automated Community Portal Site Of course, the annotation process itself has a human component: although the effort for generating the annotation of a webpage is an order of a magnitude lower than the creation of the webpage itself, there has to be some incentive to spend the extra effort. The incentive for the creation of the annotation (which is metadata for the webpage) is visibility on the web, e.g. for a *Community Web Portal*, which presents a community of interest (distributed on the web) to the outside word in a concise manner. The data collected from the annotated webpages helps automate the task of maintaining a Community Web Portal drastically. A Semantic Community Portal Website is an application demonstrating the use of ontology-based markup.

3 Related Work and Conclusion

In the Ontobroker [DEF99] and SHOE project [HHL98] means were investigated to annotate webpages with ontology-based metadata. - thus realizing part of the food chain. Agent based architectures usually focus on inter- agent communication instead to ontology creation and deployment (see [NN99] for a complaint about neglecting of the ontology problem).

We presented an information food chain, that empowers intelligent agents on the web and deploys applications that will facilitate automation of information

processing on the web. Fundamental to that approach is the use of a formal markup language for annotation of web resources.

We expect this information infrastructure to be the basis for the "Semantic Web" idea - that the World-Wide-Web (WWW) will become more than a collection of linked HTML-pages for human perusal, but will be the source of formal knowledge that can be exploited by automated agents. Without automation, and precision of operations, business and governmental uses of the information will remain limited. This food chain is partially implemented. It will be completed in the Onto-Agents-Project at Stanford University - funded in the DAML program of the DARPA.

References

[BG99] D. Brickley, R. Guha: Resource Description Framework (RDF) Schema Specification W3C Proposed Recommendation 03 March 1999, http://www.w3.org/TR/PR-rdf-schema/

[DBS98] S Decker, D. Brickley, J. Saarela, and J. Angele: A Query and Inference Service for RDF. In: Proceedings of the W3C Query Languages Workshop (QL'98), http://www.w3.org/TandS/QL/QL98/pp.html, 1998.

[DEF99] S. Decker, M. Erdmann, D. Fensel, and Rudi Studer: Ontobroker: Ontology Based Access to Distributed and Semi-Structured Unformation. In: IFIP TC2/WG2.6 Eighth Working Conference on Database Semantics (DS-8), Kluwer, 351-369, 1999.

[D98] J. Domingue: Tadzebao and webonto: Discussing, browsing, and editing ontologies on the web. In: Proc. of KAW98, Banff, Canada, 1998, http://ksi.cpsc.ucalgary.ca/KAW/KAW98/domingue/

[E97] O. Etzioni: Moving Up the Information Food Chain: Deploying Softbots on the World Wide Web, AI Magazine, 18(2): Spring 1997, 11-18.

[F98] Christiane Fellbaum (ed): Wordnet: An Electronic Lexical Database; MIT Press, 1998.

[FH97] N. Fridman Noy and C. D. Hafner: The State of the Art in Ontology Design, AI Magazine, 18(3):53-74, 1997.

[GK94] M. R. Genesereth, S.P. Ketchpel: Software Agents. In: Communications of the ACM 37 (7, July), 48-53, 1994.

[HHL98] J. Heflin, J. Hendler, and S. Luke: Reading Between the Lines: Using SHOE to Discover Implicit Knowledge from the Web. In: AAAI-98 Workshop on AI and Information Integration, 1998.

[KCT99] Peter D. Karp, Vinay K. Chaudhri, and Jerome Thomere: XOL: An XML-Based Ontology Exchange Language. ftp://smi.stanford.edu/pub/bio-ontology/xol.doc

[LS99] O. Lassila, Ralph Swick: Resource Description Framework (RDF) Model and Syntax Specification, W3C Recommendation 22 February 1999, http://www.w3.org/TR/REC-rdf-syntax/

[MGK00] P. Mitra, G. Wiederhold , and M. Kersten. A Graph-Oriented Model for Articulation of Ontology Interdependencies. In: Proceedings Conference on Extending Database Technology 2000 (EDBT'2000), Konstanz, Germany, 2000.

[NN99] H. S. Nwana, D. T. Ndumu, A Perspective on Software Agents Research, In: The Knowledge Engineering Review, Vol 14, No 2, pp 1-18, 1999.