DYNAMIC INFORMATION MANAGEMENT FOR WEB-ENABLED ENVIRONMENTS IN THE CHEMICAL PROCESS INDUSTRIES

Antonis Kokossis
University of Surrey
Guilford, Surrey GU2 7XH, United Kingdom

René Bañares-Alcántara
Univesitat Rovira i Virgili
43007 Tarragona, Spain

Abstract

We describe the development of a knowledge management platform for web-enabled environments featuring intelligence and insight capabilities. The main objective of the platform is to improve the capabilities of chemical process industries to monitor, predict and respond to technological trends and changes. The analysis, search and presentation of information retrieved from the web (or any other type of document) are explained through the use of multi-agent systems and dynamic ontologies.

The automatic evolution of dynamic ontologies requires the action of a collection of agents to extract information and discover links using classification and learning techniques. These general-purpose agents will maintain a goal to periodically access the ontology and support search functions. Conceptually similar documents would get clustered into categories and information could then be retrieved by statistical approaches or with the use of navigation patterns. Discovery of new knowledge would lead to modifications in the ontology by pruning irrelevant sections, refining its granularity and/or testing its consistency. As a whole, the system is targeted to monitor the relative impact of technologies, products and markets and update the dynamic ontology to improve decision-making.

Keywords

Knowledge Management, Ontologies, Multi-Agent Systems, Web search.

Introduction

The knowledge assets of a company consist of the knowledge regarding the products, markets, technologies and organisations of a business which enable it to add value through a set of business processes at strategic, tactical and organisational levels. In technology intensive companies the knowledge management challenges require a tentative and cautious review of the technological domains as well as venues to monitor and assess the way those domains evolve, emerge, mature, and decline. Benefits in utilising knowledge management practices include enhancement of creativity and innovation, strengthening of position, competence and increased responsiveness. The ability to grasp the dynamic profile of the discipline may translate to impressive gains of wealth and employment security. Slow reaction to developing dynamics may cost the viability of business and/or thousands of jobs. On its own, chemical engineering features a domain with hundreds of disciplines, accounts for a business sector valued over € 1.3 trillion with 34,000 enterprises only in the EU, and a domain where companies hardly maintain a national identity as they span activities all over the world.

Engineers typically assess the evolution of their disciplines by reading journals, attending conferences or, quite often, by hearsay. Instead, the web (as well as other information resources) offer scattered and distributed information that is impossible to analyse manually. It has been estimated that the World Wide Web contains more than 300 million static objects (Bharat and Broder, 1998) accessible through 100 million internet hosts (Cameron, 2002). In addition, organisations have intranets amounting to several million pages. The large majority of these documents are weakly structured. These repositories are usually searched by means of keyword-based search engines allowing a user to retrieve information by stating a combination of keywords (a query). Documents downloaded from the web are indexed according to their content, and only those matching the query (in some metric) are returned to the user. The results of this type of search usually suffer from two problems derived from the nature of the query and the lack of structure in the documents: some of the retrieved documents are irrelevant, and some of the relevant documents may not have been retrieved. While search engines provide support for the automatic retrieval of information, the tasks of extracting relevant information and its further processing remain to be done by the human user.

The performance of a search engine can be improved by the use of an ontology. In its conventional form an ontology accounts for the representation of shared concepts in a domain by specifying a hierarchy of terms facilitating communication among people (collaboration) and applications systems (integration of tools). In the case of our study, the automatic evolution of dynamic ontologies will be supported by a Multi-Agent System (MAS), i.e. a collection of agents to extract information and discover links and new concepts by using classification and learning techniques. These general-purpose agents will maintain a task to periodically access the ontology and support search functions resulting in the retrieval of documents. Conceptually similar documents would get clustered into categories; information could then be retrieved by statistical approaches or with the use of navigation patterns. Discovery of new knowledge would lead to modifications in the ontology: pruning of irrelevant sections, adding new branches, refining its granularity and/or testing its consistency.

In the next section we explain the two main techniques which form the basis of our work: ontologies and multi-agent systems. We then present the proposed architecture for a knowledge management platform and details of a case study.

Background Knowledge

Ontologies

An ontology is a vocabulary of entities, classes, properties, functions and their relationships. Ontologies are meant to provide an understanding of the static domain knowledge that facilitates knowledge sharing and reuse. As such, one can find in recent years many research efforts related to the formalisation, development and application of ontologies, most of them emerging from the Artificial Intelligence community. Fensel (2001) identifies different types of ontologies:

- Domain ontologies, e.g. for engineering, medical, modelling domains.
- Generic or Common Sense ontologies, capturing general knowledge about time, space, events, etc.
- Method ontologies, describing specific tasks, e.g. diagnosis.
- Metadata ontologies, describing the content of on-line information sources.

In contrast, there has been little work in the area of ontology building for Process Engineering applications. Perhaps the most ambitious to date is the one proposed by Batres and Naka (1999). It includes four interrelated process plant ontologies that represent the physical, behavioural and operational aspects of the plant, process and product, they are the (1) Plant Structure, (2) Material, (3) Behaviour, and (4) Management and Operation ontologies.

There are different options to represent ontologies: XML (Extended Markup Language, see http://www.xml.com), RDF (Resource Description Framework, see http://www.w3.org/RDF/), DAML+OIL (see http://www.daml.org), etc. RDF seems to be well positioned to become the standard to represent ontologies in the future (Alexaki, et al., 2000). Ontologies have been found to be useful in:

a) Retrieving the appropriate information from documents by providing a structure to annotate the contents of a document with semantical information.
b) Integrating the information from various sources by providing a structure for its organisation and facilitating the exchange of data, knowledge and models.
c) Ensuring consistency and correctness by formulating constraints on the content of information.
d) Creating libraries of interchangeable and reusable models.
e) Supporting inference to derive additional knowledge from a set of facts.

Multi-Agent Systems

An agent is an autonomous software entity which interacts proactively with its environment and with other agents in order to achieve its own goals. In a Multi-Agent System (MAS) agents exchange information in order to cooperate, negotiate and/or compete.

Agents can collect information from the web by taking advantage of semantic annotations in a document, i.e. additional information provided by the document creator. Annotations are machine processable and add structure and/or semantics to the document (meta-information). However, most of the information stored in electronic form is formatted in HTML and XML (semi structured information) or as text files (unstructured information). Wrappers can be used to extract information from semi- and unstructured documents and to parse it into a structured form, for example into a database.

The Proposed Architecture and a Case Study

We are currently developing a knowledge management platform for intelligent information retrieval in specific domains of Chemical Engineering by using a multi-agent systems approach. The information will be retrieved from both, the web and databases. The platform will integrate:

- Search agents that search the web according to some ontology.
- A search engine that allows the user to specify a query using a personalised ontology.

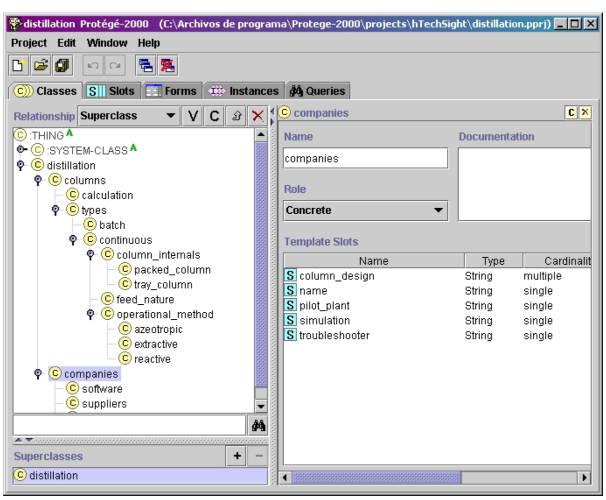

Figure 1. Definition of a distillation ontology.

- A knowledge discovery system that searches across domains (represented as ontologies) and unravels unknown dependencies.

In a first case study, being developed at URV, the ontologies are coded in RDF, a standard to include semantics into a document without making assumptions about its structure. Jena is being used to validate the ontologies structure (see http://www.hlp.hp.com/semweb/) and Protege-2000 to create and edit them (see http://protege.stanford.edu/). Figure 1 shows part of an ontology for distillation.

The multi-agent system is formed by four types of agents (Isern et al., 2002): Coordinator, Broker, Internet and User Interface agents, see Figure 2.

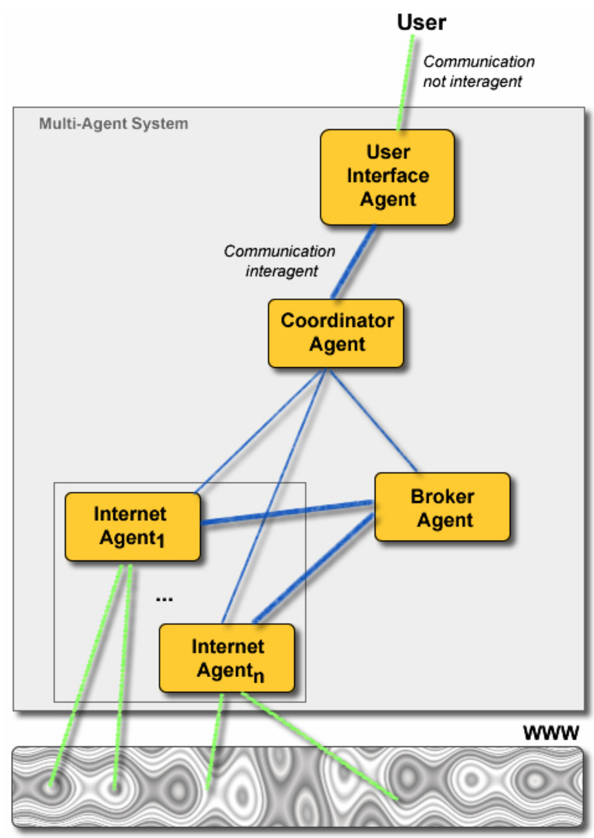

Figure 2. Architecture of the MAS.

- The User Interface agent allows the user to specify a query and to configure the initial ontology, in other words, control the target and focus of the search.
- The Coordinator agent receives the initial ontology, termed *domain ontology*, and partitions it into several *query ontologies*, assigning each one of them to a different internet agent. All the results from the search are merged by the Coordinator agent into an *information ontology*. This merging process is a complex task which takes into account the relative importance of the classes (determined by the quality of the match), eliminates redundant information, etc. In the next stage of our work new concepts will be added to the

ontology by means of incremental learning techniques.

- The Internet agents construct a knowledge base or *response ontology* containing the ordered set of documents retrieved from the web. The order can be determined using any of the classification criteria described in (Glover et al., 2002).
- The Broker agent acts as a facilitator, indicating to the Coordinator agent, for example, the state of each of the Internet agents so that the idle ones can be assigned a search task.

The flow of information through the operation of the knowledge management platform (how the different ontologies are used, transformed and created) is illustrated in Figure 3.

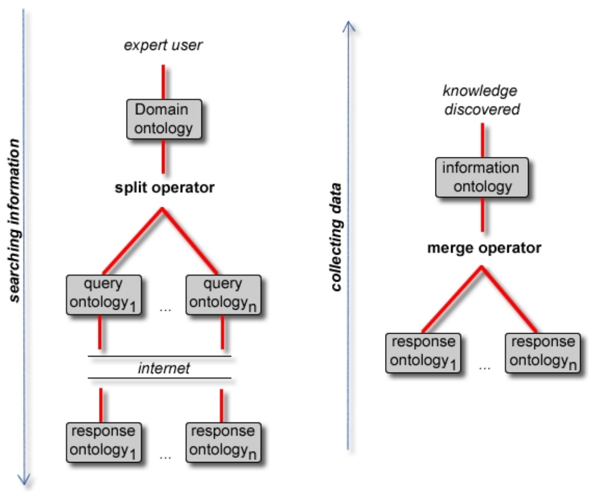

Figure 3. Information flow and the relation between ontologies.

The MAS is implemented in JADE (see http://sharon.cselt.it/projects/~jade), a set of Java classes compliant with FIPA standards for agents interoperability (see http://www.fipa.org). Thus, the agents communicate according to the protocols specified by the FIPA Agent Communicatio Language (FIPA-ACL), in particular FIPA-Query and FIPA-Request (FIPA, 2001)

Conclusions

We have described the design decisions and the initial steps for the development of a knowledge management platform for web-enabled environments featuring intelligence and insight capabilities. The main objective of the platform is to improve the capabilities of chemical process industries to monitor, predict and respond to technological, product and market trends and changes. A company could assess changes in the market, competitor's profile, available technologies, and evaluate the penetration of new products and services and new areas of growth. The retrieval, analysis, modification and presentation of information retrieved from the web (or any other type of resource) are achieved through the use of multi-agent systems and dynamic ontologies.

Ontologies are used to specify the queries and the search is done by a set of Internet agents each of them acting on a part of the original ontology. The retrieved information is organised and merged into a single knowledge base where discovery of new knowledge has taken place by pruning irrelevant sections of the original ontology, refining its granularity and/or testing its consistency.

Acknowledgements

This work has been funded by the EU project hTechSight, IST-2001-33174.

References

Alexaki, S., et al. (2000). Managing RDF Metadata for Community Webs. *Proceedings 2nd Int. Workshop on the WWW and Conceptual Modelling (WCM'2000)*, Salt Lake City, USA, 140-151.

Bharat, K. and A. Broder (1998). A technique for measuring the relative size and overlap of public Web search engine. *Proceedings of the 7th Int. WWW Conf.*, Brisbane, Australia.

Cameron, I. (2002). Web-based CAPE Systems – now and the future – *CAPE Forum*, Tarragona, Spain.

Fensel, D. (2001). Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce. *Springer*, Heidelberg, Germany.

FIPA, (2001). FIPA Communicative Act Library Specification, FIPA Specification Document XC00037J, available at http://www.fipa.org.

Glover, E.J., et al. (2002). Using Web Structure for Classifying and Describing Web Pages, *Proceedings WWW2002*, Honolulu, USA.

Isern, D., D. Riaño and A. Aldea. (2002). Cerca d'informació basada en coneixement usant un sistema multi-agent. Submitted to *CCIA 2002*, Spain.