Community is Knowledge! in $(KA)^2$

V. Richard Benjamins[1,3] and Dieter Fensel[2]

[1] Artificial Intelligence Research Institute (IIIA), Spanish Council for Scientific Research (CSIC), Campus UAB, 08193 Bellaterra, Barcelona, Spain, richard@iiia.csic.es, http://www.iiia.csic.es/~richard

[2] University of Karlsruhe, Institute AIFB, 76128 Karlsruhe, Germany, dfe@aifb.uni-karlsruhe.de, http://www.aifb.uni-karlsruhe.de/WBS/dfe/

[3] Dept. of Social Science Informatics (SWI), University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands, richard@swi.psy.uva.nl, http://www.swi.psy.uva.nl/usr/richard/home.html

Abstract

The **K***nowledge* **A***nnotation Initiative of the* **K***nowledge* **A***cquisition Community,* *(KA)*^{2} is an initiative to develop an ontology that models the knowledge acquisition community (its researchers, topics, products, etc.). This ontology will form the basis to annotate WWW documents of the KA community in order to enable intelligent access to these documents. $(KA)^2$ is an open joint-initiative where the participants are actively involved in (i) a distributive ontological engineering process to model the knowledge acquisition community (a domain ontology), and (ii) annotating web pages relevant for the KA community (the instances of the domain ontology).

$(KA)^2$ aims at intelligent knowledge retrieval from the Web and automatic derivation of "new" knowledge. In other words, it aims at knowledge-based reasoning on the Web, as opposed to information retrieval. Another objective of the initiative is to get better insight in distributive ontological engineering processes.

1 Introduction and motivation

The $(KA)^2$ initiative[4] has three major motivations and contributions. First, the World-Wide Web can be seen as the largest knowledge base ever (even bigger than CYC (Lenat and Guha, 1990)). However, the amount of inferencing and deduction of new knowledge on the WWW is very limited. Current search engines (like Altavista or Yahoo) are mostly keyword based and basically do *information retrieval*. This leads, as everybody might have experienced, to answers containing overwhelming amounts of references to web documents. Although search engines get increasingly smarter, we expect that there will be a limit to such keyword-based information retrieval. An alternative approach concerns so-called *ontology-based* knowledge access or retrieval. An ontology refers to a commonly agreed conceptualisation of some domain. One of the issues $(KA)^2$ aims to investigate, is the power and role of ontologies in intelligent access to information on the Web. In this sense, $(KA)^2$ hopes to contribute to the solution of a significant problem.

A second motivation of the $(KA)^2$ initiative relates to ontological engineering. Ontologies attract nowadays much attention of a variety of research communities (Guarino and Poli, 1995), illustrating the fact that ontologies are considered useful for many applications. The notion of ontology, however, has been somewhat diluted lately. Many

[4]The URL of the $(KA)^2$ homepage is `http://www.aifb.uni-karlsruhe.de/WBS/broker/KA2.html`

specific domain models (e.g. taxonomies) are currently called ontologies, regardless of the fact that these ontologies might only reflect the opinion of one or several persons, and basically only contain classes and sub-classes (and no axioms). Building a *consensual* and *rich* ontology is, however, not an easy task as it requires agreement of different people on different aspects. Concerning the KA ontology for example, in the Dutch university system, a Ph.D. student can officially only be supervised by a full professor, which would give rise to the ontological axiom: If X is supervisor of Y and Y is a Ph.D. student, then X is a full professor. In Spain, on the other hand, a Ph.D. student can be supervised by either a full professor or a doctor, making the axiom above invalid. $(KA)^2$ is an international initiative whose aim is to build a consensual ontology in a distributive way. A contribution of $(KA)^2$ is that it can be viewed as a large-scale experiment in collaborative, distributive ontology construction.

A third motivation of $(KA)^2$ is to have a clear insight in the groups and topics of the knowledge acquisition research community. To come up with a commonly agreed conceptualisation and classification of the work and the people active in the KA community, is an important contribution in itself. Moreover, if this knowledge is easily and intelligently accessible, it could be very helpful to stimulate cooperations between different groups, to unite forces and to prevent repetitions of work.

The structure of this paper is as follows. In Section 2, we mention a disclaimer of our initiative, restricting its scope for feasibility reasons. In Section 3, we describe the approach to achieve the initiative's objectives. Section 4 discusses Ontobroker, which includes an ontology-based web-crawler. In Section 5, we briefly sketch the organisational structure of the initiative. Finally, in Section 6 we conclude the paper.

2 Feasibility of the approach and deliberative restrictions

As outlined above, one of the objectives of $(KA)^2$ is to turn the WWW from a knowledge base into a knowledge-based system, using an ontology and by developing an interpreter. However, it is infeasible and unthinkable (and even undesirable) that the whole World-Wide Web would agree on one unique ontology. This would imply that all people shared the same view on the world. Nothing is less true.

Therefore, we used the metaphor of a newsgroup: a group of people that share a common subject and a related point of view on this subject (Fensel et al., 1997a). This allows people – we call them an *ontogroup* – to annotate their web pages based on a shared ontology[5] to enable automatic inference.

In $(KA)^2$, we are defining such an ontogroup as the knowledge acquisition community. The web sites of the KA community form a sub-web of the WWW, and we think it is feasible to come up – in a *distributive* and *collaborative* way – with a KA-community accepted view on the KA world.

[5] Actually, "shared ontology" is a pleonasm since an ontology is by definition shared. We write it here to stress the consensus aspect of an ontology, which is not always a characteristics of existing ontologies.

3 The approach

There are three main issues involved in the initiative (see Figure 1). (i) The knowledge acquisition community has to built its own ontology. (ii) The community has to fill this ontology with instances by annotating the relevant web pages. (iii) Given a query, a web-crawler has to access the web pages and use the ontology to provide answers. Depending on how rich the ontology is (e.g. the amount axioms allowing inferencing), the web-crawler can also deduce "new" information that is not explicitly stored on the Web. Notice that such inferencing is very common in knowledge-based systems, but not at all for web search engines.

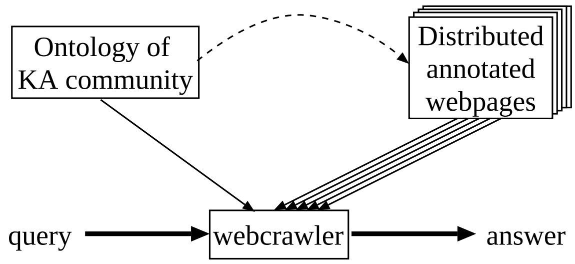

FIGURE 1: Overview of the $(KA)^2$ initiative. Relevant web pages of the knowledge acquisition community are annotated in terms of the KA ontology (dotted, bowed arrow). A web-crawler gets information from the ontology and from the web pages (the instances of the ontology) and based on that, it can deduce "intelligent" answers to queries.

3.1 Distributive ontological engineering

In order to come up with a consensual ontology of the knowledge acquisition community, we build the ontology as a collaborative joint-effort of the whole KA community. This requires that the ontology can be easily inspected, browsed and downloaded. These requirements have lead us to use the Ontolingua server (Farquhar et al., 1996). Ontolingua is an interactive environment especially useful for updating, maintaining and browser ontologies. Ontolingua ontologies can be translated to several different languages, including Prolog, CORBA's IDL (Orfali et al., 1996), CLIPS, LOOM (MacGregor, 1991), KIF, Epikit (Genesereth, 1992).

The current version of the ontology can be viewed at the European mirror site in Madrid of the Ontolingua server of Stanford University (http://www-ksl-svc-lia.dia.fi.upm.es:5915/). Login as "ontologias-ka2" with password "adieu007". The ontology for the KA community consists of eight related ontologies: an organisation ontology, a project ontology, a person ontology, a research-topic ontology, a publication ontology, an event ontology, a research-product ontology and a research-group ontology.

Ontological primitives Ontologies built in Ontolingua use the Frame Ontology (Gruber, 1993), which is written in KIF (Knowledge Interchange Format) (Genesereth and Fikes, 1992). The Frame Ontology is, as its name suggests, a frame-based language which includes primitives such as classes, sub-classes, attributes, values, relations and axioms. Related ontologies can be connected to each other by inclusion.

Design decisions Before starting to build the (current version of the) ontology of the KA community, we took several ontological design decisions in line with the goal of the $(KA)^2$ initiative.

- The ontology should be simple. If we want to have maximal participation of the active research groups in the KA community, then using the ontology should be easy and straightforward, otherwise there is the danger to get lost in details. It should be – as much as possible – unambiguous for the provider agents to model particular instances (provider agents provide the initiative with knowledge by annotating their web pages, see Section 5). This made us decide to not use some of the already existing ontologies in Ontolingua, but rather to start from scratch focusing only on the most relevant concepts. Later versions of the ontology can be more extensive.

- The ontology should be modular and allow reusability. The ontology is currently divided in eight separate, but closely related, ontologies. Each of these can be reused in other ontologies. In particular, several of these sub-ontologies are not specific for the KA community, but could be used for modelling any research community. Examples are the ontologies for "person", "publication" and "organisation".

- The ontology should have high "visibility" and be easy accessible. This made us take the decision to use the Ontolingua server.

A distributive joint-effort Building the ontology is a collaborative and distributed process of the KA community. So-called ontopic agents (from ontology topic) can construct parts of the ontology about which they have profound knowledge. For example, if some research group works on "verification and validation", then that group could suggest a sub-ontology of the research-topic ontology about V&V. Ontolingua comes with an "ontology editor" that allows developers to input classes, sub-classes, attributes, values, axioms, etc. in a structured way, and the editor automatically generates Ontolingua code.

Although the ontology editor helps, many people may have experienced that building an ontology from scratch in Ontolingua is daunting, not in the last place because of slow network connections. Experience has shown that the Ontolingua editor is better suited for checking, maintaining and modifying the ontology than for building an ontology from scratch. Therefore, an alternative strategy is to build ontologies off line, and then import them into Ontolingua. However, writing Ontolingua code is not a comfortable level for persons to work with, that is, it is too close to the symbol level. To overcome this problem, ODE (Gomez-Perez et al., 1996) has been developed (Ontological Design Environment) and it allows developers to specify their ontology at a conceptual level by means of completing tables (see Table 1). These tables are then automatically translated into Ontolingua code, which can be included in the ontology at the Ontolingua server.

Representation of the ontology As will be presented in Section 4, our web-crawler reasons with Frame Logic (FLogic). This means that the Ontolingua ontology also has to be available in FLogic. We deliberatively did not choose for doing the collaborative ontological engineering process in FLogic for two reasons. (i) Ontolingua comes with an integrated environment to develop ontologies, which is not the case for FLogic. (ii) Ontolingua is well known, which enhances the visibility of the ontology and of the initiative.

Concept name	Synonym	Description	Instances	Attributes
Person				Photo First-name Last-name Address Email ...
Researcher			Nicola Guarino Enric Plaza Tom Gruber Rose Dieng Henrik Eriksson Nigel Shadbolt ...	Research-interest Member-of Cooperates-with ...
PhD-Student			Mariano Fernandez Lopez Stefan Decker	Has-Supervisor
...	...	...	...	...

TABLE 1: Using tables to specify an ontology. A small part of how to specify the "person-ontology".

However, a consequence of this decision is that we have to provide translators to establish a formal connection between the two. Basically there are two possibilities. (1) Translators from ODE to both Ontolingua and FLogic. Equivalence between the two is guaranteed by always modifying the ontology in ODE. (2) A translator from Ontolingua to FLogic. If in addition a translator from FLogic to Ontolingua is built, then it becomes also possible to inspect the instances if the ontology at the Ontolingua server. Notice that the current instances of the ontology have been entered manually, but in the course of the initiative they will be collected from the distributed web pages of the KA community.

Examples of the ontology As mentioned above, the KA ontology currently comprises eight different ontologies (about organisations, projects, persons, research-topics, publications, events, research-products and research-groups). We have to stress that these represent the current version of the ontology. It is the aim of $(KA)^2$ to come up with a consensual version. In the following, we show global overviews of three sub-ontologies: the Research-product ontology, the Person-ontology and the Publication-ontology.

The Research-product ontology defines the products that the KA community develops such as all kinds of knowledge elicitation tools, validators, modelling languages, etc. The two relations model that a product is "developed-by" a research group and "produced-by" a project. The Research-product ontology comprises 19 classes and 2 relations. The following overviews do not show which classes the relations connect (but it can be browsed in Ontolingua).

```
Class hierarchy (19 classes defined):

    Product
       Computer-Support
          Elicitation-Tool
```

```
        Implementation-Environment
        Intelligent-Editor
        Internet-Tool
        Natural-Language-Parser
        Ontology-Library
        Problem-Solving-Method-Library
        Transformational-Tool
        Validator
        Verifier
     Knowledge-Acquisition-Methodology
        Guideline
        Modeling-Language
        Ontology
        Paper-Library
        Problem-Solving-Method
        Specification-Language

2 relations defined:

  Developed-By
  Produced-By
```

The Person-ontology defines the types of persons working in academic environments, along with their characteristics. This ontology defines 10 classes and 23 relations.

```
Class hierarchy (10 classes defined):

    Person
       Employee
          Academic-Staff
             Lecturer
             Researcher
          Administrative-Staff
             Secretary
             Technical-Staff
       Student
          Phd-Student

23 relations defined:

  Address
  Affiliation
  Cooperates-With
  Editor-Of
  Email
  First-Name
  Has-Publication
  Head-Of-Group
  Head-Of-Project
  Last-Name
  Member-Of-Organization
  Member-Of-Program-Committee
  Member-Of-Research-Group
  Middle-Initial
  Organizer-Of-Chair-Of
  Person-Name
```

```
Photo
Research-Interest
Secretary-Of
Studies-At
Supervises
Supervisor
Works-At-Project
```

The Publication-ontology defines – in 13 classes and 28 relations – the usual bibliographic entities and attributes. We tried, however, to keep it manageable.

```
Class hierarchy (13 classes defined):

    On-Line-Publication
    Publication
       Article
          Article-In-Book
          Conference-Paper
          Journal-Article
          Technical-Report
          Workshop-Paper
       Book
       Journal
          IEEE-Expert
          IJHCS
          Special-Issue

28 relations defined:

  Abstract
  Book-Editor
  Conference-Proceedings-Title
  Contains-Article-In-Book
  Contains-Article-In-Journal
  Describes-Project
  First-Page
  Has-Author
  Has-Publisher
  In-Book
  In-Conference
  In-Journal
  In-Organization
  In-Workshop
  Journal-Editor
  Journal-Number
  Journal-Publisher
  Journal-Year
  Last-Page
  On-Line-Version
  On-Line-Version-Of
  Publication-Title
  Publication-Year
  Technical-Report-Number
  Technical-Report-Series
  Type
  Volume
  Workshop-Proceedings-Title
```

3.2 The annotation language

The problem with information retrieval from the Web is that there is no commonly used syntax for representing semantics. Current search engines are therefore restricted to keyword-based search, and retrieve information by syntactically matching input words with words appearing in web documents. This “keywordness” is the reason for the overwhelming amount of (also) irrelevant answers on a query.

Basically, the cause of the problem is that HTML does not allow to specify semantics. HTML is only concerned with marking up text and providing hyperlinks, but not with the *content* of the information. There are some approaches to extend HTML with meta-information to state things about the content of web documents, such as the “meta” tag, Guha’s Meta Content Format (Guha, 1996), and the Extensible Markup Language (XTM) (http://www.w3.org/XML/). These approaches extend HTML in the sense that they provide a means to express information *about* a web document. However, they do not tell you what knowledge to include.

The (KA)^2 initiative is in particular concerned with *what* knowledge to represent, and its approach is based on ontologies. However, in order to express ontological information in web pages or documents (written in HTML), some syntax is needed. For the purpose of (KA)^2 it suffices to simply add one new attribute to the anchor tag of HTML: the *onto* attribute. This attribute does not affect the visualisation of HTML documents by standard web browsers such as Netscape or Explorer. The only thing that the onto attribute does, is that it makes visible valuable pieces of knowledge for the web-crawler – in the same way as (only) glittering objects in the world are visible for a crow. This small extension of HTML has been chosen to keep annotation as simple as possible to lower the threshold for participants of the initiative. Also, it enables the direct usage (actually, reuse) of textual knowledge already in the body of the anchor, as well as of further information provided by the other anchor attributes. This prevents the knowledge provider from representing the same piece of information twice. In our case, this simple solution suffices because only *factual* ontological knowledge is contained in HTML pages (Fensel et al., 1998). If, in the future, the need arises to include more elaborate ontological knowledge in web documents, we may develop an extension of HTML according to the lines of XML.

Figure 2 illustrates fragments of an example web page annotated with the `onto` attribute. For example, `page` in ⟨`a ONTO="page[address=body]"`⟩ refers to the URL of the web page. `Body` refers to what follows and what is within the scope of the anchor, i.e. until the closing ⟨`/a`⟩. `Address` is a class of the KA ontology. In general, all values of the onto attribute should come from the KA ontology.

4 Ontobroker

Having discussed the KA ontology and the annotated web pages, in this section, we will present a brokering service that uses that knowledge to make intelligent deduction. The ontology-based brokering service Ontobroker[6] consists of three main elements: a web-crawler (called Ontocrawler), an inference engine and a query interface. Each of these

[6]The URL of Ontobroker is `http://www.aifb.uni-karlsruhe.de/WBS/broker/`

```
---------------------------------------------------------------------
<html>
<head><TITLE> Richard Benjamins </TITLE>
<a ONTO="page:Researcher"> </a>
</head>

<H1> <A HREF="pictures/id-rich.gif">
<IMG  align=middle SRC="pictures/richard.gif"></A>
<a ONTO="page[photo=href]"
HREF="http://www.iiia.csic.es/~richard/pictures/richard.gif" ></a>
<a ONTO="page[firstName=body]">Richard</a>
<a ONTO="page[lastName=body]">Benjamins </a>
</h1> <p>

<A ONTO="page[affiliation=body]" HREF="#card">
Artificial Intelligence Research Institute (IIIA)</A> -
<a href="http://www.csic.es/">CSIC</a>, Barcelona, Spain <br>
and <br>
<A ONTO="page[affiliation=body]" HREF="http://www.swi.psy.uva.nl/">
Dept. of Social Science Informatics (SWI)</A>
-
<A HREF="http://www.uva.nl/uva/english/">UvA</A>, Amsterdam, the
Netherlands

<DL>
<DT><STRONG><A HREF="../../IIIA.html">IIIA</A> -
<a ONTO="page[address=body]">
Artificial Intelligence Research Institute </STRONG>
<DT><EM>CSIC - Spanish Scientific Research Council</EM>
<DT>Campus UAB
<DT>08193 Bellaterra, Barcelona, Spain </a>
<DT><IMG SRC="gifs/tel.gif">
    voice: +34-3-580 95 70
<DT><IMG SRC="gifs/fax.gif">
    fax: +34-3-580 96 61
<DT><IMG SRC="gifs/email.gif">
    Email:<A HREF="mailto:richard@iiia.csic.es" ONTO="page[email=href]">
          richard@iiia.csic.es</A>
<DT>URL: <A HREF="http://www.iiia.csic.es/~richard/">
         http://www.iiia.csic.es/~richard</A>
</DL></font>

</body>
</html>
---------------------------------------------------------------------
```

FIGURE 2: Example web page annotated with the ONTO attribute. `Page` in ⟨`a ONTO="page[address=body]"`⟩ refers to the URL of the page. `Body` refers to what follows and what is within the scope of the anchor, i.e. until the closing ⟨`/a`⟩. `Address` is a class of the KA ontology.

elements is accompanied by a formalisation language: the annotation language for annotating web documents with ontological information, the representation language for specifying ontologies (inside Ontobroker), and the query language for formulating queries. Notice that, although we use Ontobroker for $(KA)^2$, it is not specific for this initiative. Given any ontology and correspondingly annotated web pages, Ontobroker can deliver its brokering service.

Ontocrawler First, Ontocrawler searches through a fragment of the WWW that is annotated – using the annotation language – according to a particular ontology (in our case, the KA ontology) and collects the annotated knowledge fragments. Second, it realises a wrapper that translates annotated web documents into facts formulated in the representation language. Neither the inference engine nor the querying client have to be aware of the syntactical way, the facts are represented on the web in the annotation language. Ontocrawler provides this abstraction mechanism. Only the knowledge provider has to use the annotation language.

In order to become a provider of an ontologically annotated knowledge chunk on the WWW, one has to do two things:

1. Define an O-page and register the page's URL at the provider index of Ontocrawler. This O-page contains a sub-index that specifies all URLs of the annotated web documents of that provider. Figure 3 gives an example.
2. Use the annotation language and an ontology of the Ontobroker to annotate the concerned web documents.

```
-----------------------------------------------------------
http://www.iiia.csic.es/~richard/index.html
http://www.iiia.csic.es/~richard/activities.html
http://www.iiia.csic.es/~richard/interests.html
http://www.iiia.csic.es/~richard/projects.html
http://www.iiia.csic.es/~richard/publications/pub-type.html
http://www.iiia.csic.es/~richard/cv/cv.html
-----------------------------------------------------------
```

FIGURE 3: An O-page of a knowledge provider agent.

Inference engine The inference engine receives the query of a client and uses two information sources for deriving an answer: the ontology chosen by the client (the KA ontology, in our case) and the facts that were found by Ontocrawler on the WWW. The basic inference mechanism of the inference engine is the derivation of a minimal model of a set of Horn clauses (see (Fensel et al., 1998) for more details). However, the language for representing ontologies is syntactically enriched. First, ideas of (Lloyd and Topor, 1984) were used to get rid of some of the limitations of Horn Logic, without requiring a new inference mechanism. Second, languages with richer epistemological primitives than predicate logic are provided. Frame logic (Kifer et al., 1995) is used as the representation language for ontologies inside Ontobroker. It incorporates objects, relations, attributes, classes, and is-subclass-of and is-element-of relationships within a first-order semantic framework.

Query interface The broker has to communicate with clients who ask for some knowledge using web browsers like Netscape and Explorer. The query interface of Ontobroker is realised through a couple of active HTML pages and cgi-scripts that are executed by the browser of the client. The client selects the KA ontology to formulate his query. The answer of the broker will be based on this ontology and on the web documents that have been annotated using this ontology (only if an O-page has been registered, of course). The query language is a subset of the representation language customised for formulating queries.

The query formalism is oriented towards Frame-Logic syntax, that defines the notion of instances, classes, attributes and values. The generic schema for this is `O:C[A--⟩⟩V]` meaning that the object O is an instance of the class C with an attribute A that has a certain value V. At each position in the above schema variables, constants or arbitrary expressions can be used. In the following we will provide some example queries to illustrate our approach.

```
FORALL R <- R:Researcher.
```

This query asks for all known objects, which are instances of the class researcher. Because the object identifier of a researcher is his/her homepage-URL, this query would result in a large list of URLs. This is one of the simplest possible queries. However, usually we are not interested in all researchers, instead we are interested in information about researchers with certain properties, e.g., we want to know the homepage, the last name and the email address of all researchers with first name "Richard". To achieve this we can use the following query:

```
FORALL Obj, LN, EM <- Obj:Researcher[firstName->>"Richard";
lastName->>LN; email->>EM].
```

The Ontobroker gives the following answer (actually, there is only one researcher in the knowledge base whose first name is "Richard"):

```
Obj = "http://www.iiia.csic.es/~richard/index.html"
LN = "Benjamins"
EM = "mailto:richard@iiia.csic.es"
```

Another possibility is to query the knowledge base for information about the ontology itself, e.g. the query:

```
FORALL Att, T <- Researcher[Att=>>T]
```

asks for all attributes of the class Researcher and their associated classes. Figure 4 shows part of the answer of Ontocrawler. At the top left, the client has chosen to query the Knowledge Acquisition community. A bit lower, one can see the query, and below that the answer of Ontobroker appears (`Att` denotes "attribute" and `T` the type of the value of the attribute).

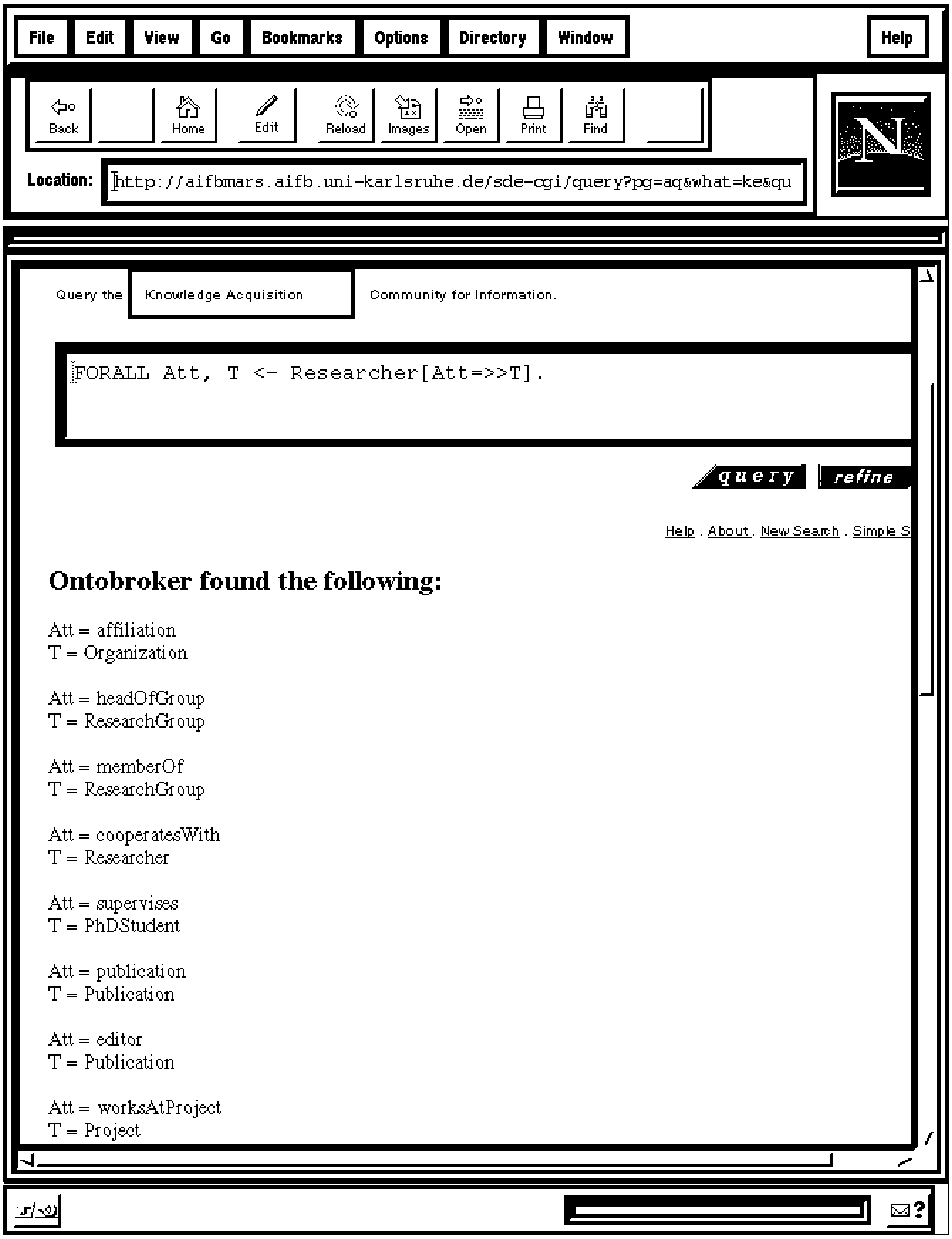

FIGURE 4: Ontobroker in action.

5 Organisation of $(KA)^2$

$(KA)^2$ is organised as a community of several types of agents. Each type has well-defined responsibilities in order to get the $(KA)^2$ initiative started, keep it going, assure its scientific content, make it a global collaborative effort and attract industrial interest: coordinating agents, provider agents, ontopic agents, wise agents and business agents.

Coordinating agents The coordinating agents are responsible for the daily matters of the initiative. There are 6 of these agents. The ontology agent (Asuncion Gomez-Perez, LIA, UPM) is responsible for keeping the KA ontology always up-to-date at the Ontolingua server. The Webtool agent (Enrico Motta, KMI, OU) takes care of the web issues involved in the communication between the agents such as setting up a mailing list and a mail archive, as well as providing web tools to collaboratively work on the same ontology. The managing agent (Richard Benjamins, IIIA-CSIC, SWI-UvA) is responsible for the collaborative ontological engineering process for building the KA ontology, and for the overall process of the initiative. The recruiting agent (Dieter Fensel, AIFB, UKa) tries to convince KA groups to participate in the initiative (he might make you an offer you can't refuse). The annotation agent (Michael Erdmann, AIFB, UKa) coordinates the process of annotating web pages, and the ontobroker agent (Stefan Decker, AIFB, UKa) is responsible for keeping the Ontobroker up and working. Finally, the "window on USA" agent (Mark Musen, Stanford, SMI) informs the initiative on related events, initiatives and work in the USA.

Wise agents Wise agents are concerned with the scientific issues involved in the initiative. They give high-level steering and suggestions concerning whether the initiative is going into the right direction. Wise agents currently include Bob Wielinga (Univ. Amsterdam, the Netherlands), B. Chandrasekaran (Ohio State Univ., USA), Rudi Studer (AIFB, Univ. Karlsruhe, Germany), Bill Swartout (ISI, Univ. Southern California, USA), James Hendler (Univ. Maryland, USA), Brian Gaines (Univ. Calgary, Canada).

Provider agents Provider agents provide the initiative with instances of the ontology. In other words, they have to annotate their web pages. At the kick-off meeting during EKAW'97, the following groups and people committed themselves to be a provider agent. The recruiting agent is responsible for attracting more researchers and groups.

Provider agents: Andreas Abecker (DFKI, Germany), Nathalie Aussenac (IRIT, Univ. Paul Sabatier, France), Maillet-Contoz (LIRMM, France), Sean Wallis (Univ. College London, United Kingdom), Robin Boswell, Susan Craw (Robert Gordon Univ., United Kingdom), Enrico Motta KMI, (Open Univ. United Kingdom), Enric Plaza, Richard Benjamins (IIIA-CSIC, Spain), Christine Pierret (Univ. Rennes, France), Dieter Fensel, Rudi Studer, Michael Erdmann, Stefan Decker (AIFB, Univ. Karlsruhe, Germany), Asuncion Gomez (Technical University of Madrid, UPM, Spain), Bob Wielinga, Richard Benjamins (SWI, Univ. of Amsterdam, the Netherlands), Hans Akkermans (Univ. Twente, the Netherlands), B. Chandrasekaran (Ohio State Univ., USA), Derek Sleeman (Univ. of Aberdeen, United Kingdom), Nigel Shadbolt (Univ. of Nottingham, United Kingdom), Paul Compton, Tim Menzies (University of New South Wales, Australia), Frances Brazier, Niek Wijngaards, Frank van Harmelen, Annette ten Teije (Free Univ. Amsterdam, the Netherlands).

Ontopic agents Ontopic agents are research groups that contribute to the ontological engineering process to establish a consensual ontology of the KA community. There will be about 15 groups of ontopic agents, each group being responsible for a particular topic of the KA ontology. The Webtool agent will provide ontopic agents with webtools to distributively construct (parts of) the ontology.

Business agents Business agents are responsible for exploring the possibility of external funding of the initiative and raising the interest of possible interested industries. Currently, there is one business agent: Annejet Meijler of the Intelligent Systems Lab Amsterdam of the University of Amsterdam, the Netherlands.

6 Conclusions

In this paper, we presented an initiative – $(KA)^2$ – whose goal is to enable knowledge-based reasoning on (a subpart of) the WWW, using an ontology. The subpart concerns the web pages of the KA community, and many research groups and researchers are already involved. To achieve the objectives of $(KA)^2$ three things are needed: (1) an ontology of the KA community, (2) annotated web pages in terms of the ontology, and (3) an ontology-based web-crawler to perform reasoning. Constructing the KA ontology will be a collaborative and distributed process for which the Ontolingua server has been chosen. The instances of the ontology are provided distributively by KA researchers through annotating their relevant web pages.

The idea of using ontologies to annotate information on the WWW is also part of the SHOE-approach (Luke et al., 1996; Luke et al., 1997). HTML pages are annotated via ontologies to support information retrieval based on semantic information. However, there is a main differences in the underlying philosophy. Providers of information in SHOE can introduce arbitrary extensions of ontologies and no central provider index is defined. As a consequence, the client may not know the ontological terms that he must use in a query and the web crawler may miss knowledge chunks because it cannot parse the entire WWW. In SHOE, ontologies are proposed as gradual improvement of the competence of global search engines on the WWW. If the user knows – for some reasons – parts of the ontology (like he has to know the right key words) and if the search engines knows – for some reasons – the appropriate URLs (for example, by executing keyword search on ontological terms), then it can be used for a semantically guided search through the web. Our approach is based on a joint ontological engineering activity of a group of web users that establish a consensual point of view. As a consequence we can provide the entire ontology used for annotation to the questioner and we can deliver complete answers. This ontology may be useful also for different purposes besides their application to the web. Finally, we extend the search metaphor of SHOE to the capability to express complex inferences using the knowledge as it is provided by the web. The ontological formalism used by SHOE is rather limited in regard to this purpose. Technically, the main difference stems from the fact that SHOE uses description logic whereas Ontobroker relies on Frame-Logic (a deductive object oriented database language). Precise comparisons of both representation and reasoning paradigms are still ongoing research activities (Kandzia and Schlepphorst, 1996; Fensel et al., 1997b).

One of the objectives of $(KA)^2$ is to investigate is the power and role of ontologies in intelligent access to information on the Web. We therefore think that applying these ideas in an industrial or commercial setting could be interesting. To stay close to the $(KA)^2$ initiative, think for example about the usefulness of such knowledge-based reasoning capabilities for scientific publishers like Elsevier-Kluwer, Academic Press, Addison Wesley, etc. In general, the potential advantages of more intelligent reasoning on the WWW are enormous.

The current status of the $(KA)^2$ initiative is that all provider agents have to annotate their web pages using the ontology. However, using Machine Learning techniques it should be possible to automatically learn the instances from the web pages using the KA ontology as background knowledge. In a more distant future, it may also become possible to learn, derive or mine (parts of) the ontology (semi) automatically. For instance, statistical and ML techniques could be used to identify the most frequently occurring concepts at pages of the KA community, and try to cluster them. These clusters could then suggest a basic structure or starting point for the ontology. This is not so much of interest for our current initiative, but it is extremely valuable if our initiative shows that ontology-based knowledge retrieval and reasoning is a good alternative for keyword-based information retrieval. In general, it is undoable to build large ontologies as a collaborative process as we do for $(KA)^2$. In our initiative, however, it is worth the effort because we are still investigating the role of ontologies on the Internet.

Last but not least, the ontobroker agent is currently improving the query interface of Ontobroker so that clients do not have to write their queries in FLogic.

Acknowledgement

We would like to thank all participants of EKAW'97 for their input. We acknowledge Enric Plaza for the slogan title. We thank Asunción Gomez-Perez for her comments on the paper. Richard Benjamins is supported by the Netherlands Computer Science Research Foundation with financial support from the Netherlands Organisation for Scientific Research (NWO), and by the European Commission through a Marie Curie Research Grant (TMR).

References

[Farquhar et al., 1996]Farquhar, A., Fikes, R., and Rice, J. (1996). The ontolingua server: A tool for collaborative ontology construction. In Gaines, B. R. and Musen, M. A., editors, *Proceedings of the 10th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop*, pages 44.1–44.19, Alberta, Canada. SRDG Publications, University of Calgary. http://ksi.cpsc.ucalgary.ca:80/KAW/KAW96/KAW96Proc.html.

[Fensel et al., 1998]Fensel, D., Decker, S., Erdmann, M., and Studer, R. (1998). Ontobroker: transforming the WWW into a knowledge base. In *KAW'98*, page Submitted.

[Fensel et al., 1997a]Fensel, D., Erdmann, M., and Studer, R. (1997a). Ontology groups: Semantically enriched subnets of the www. In *Proceedings of the International Workshop Intelligent In-*

formation Integration during the 21st German Annual Conference on Artificial Intelligence, Freiburg.

[Fensel et al., 1997b]Fensel, D., Rousset, M.-C., and Decker, S. (1997b). Workshop on comparing description and frame logics. *Data and Knowledge Engineering*, to appear.

[Genesereth, 1992]Genesereth, M. R., editor (1992). *The Epikit manual.* Epistmemics, Inc, Palo Alto, CA.

[Genesereth and Fikes, 1992]Genesereth, M. R. and Fikes, R. E. (1992). Knowledge interchange format, version 3.0, reference manual. Technical report, Logic-92-1, Computer Science Dept., Stanford University. http://www.cs.umbc.edu/kse/.

[Gomez-Perez et al., 1996]Gomez-Perez, A., Fernandez, A., and Vicente, M. D. (1996). Towards a method to conceptualize domain ontologies. In *Working notes of the workshop on Ontological Engineering, ECAI'96*, pages 41–52. ECCAI.

[Gruber, 1993]Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5:199–220.

[Guarino and Poli, 1995]Guarino, N. and Poli, R. (1995). The role of ontology in the information technology. *International Journal of Human-Computer Studies*, 43(5/6):623–965. Special issue on ontology.

[Guha, 1996]Guha, R. V. (1996). Meta content framework: A white paper. Technical Report http://mcf.research.apple.com/wp.html, Apple Research.

[Kandzia and Schlepphorst, 1996]Kandzia, P.-T. and Schlepphorst, C. (1996). DOOD and DL - do we need an integration. In *Proceedings of the Workshop on Internet-based Information Systems, AAAI'96.* http://www.cs.umd.edu/projects/plus/SHOE/.

[Kifer et al., 1995]Kifer, M., Lausen, G., and Wu, J. (1995). Logical foundations of object-oriented and frame-based languages. *Journal of the ACM.*

[Lenat and Guha, 1990]Lenat, D. B. and Guha, R. V. (1990). *Representation and Inference in the Cyc Project.* Addison-Wesley.

[Lloyd and Topor, 1984]Lloyd, J. W. and Topor, R. W. (1984). Making prolog more expressive. *Journal of Logic Programming*, 3:225–240.

[Luke et al., 1996]Luke, S., Spector, L., and Rager, D. (1996). Ontology-based knowledge discovery on the world-wide web. In *Proceedings of the Workshop on Internet-based Information Systems, AAAI'96.* http://www.cs.umd.edu/projects/plus/SHOE/.

[Luke et al., 1997]Luke, S., Spector, L., Rager, D., and Hendler, J. (1997). Ontology-based web agents. In *Proceedings of First International Conference on Autonomous Agents.* http://www.cs.umd.edu/projects/plus/SHOE/.

[MacGregor, 1991]MacGregor, R. (1991). Inside the LOOM classifier. *SIGART Bulletin*, 2(3):70–76.

[Orfali et al., 1996]Orfali, R., Harkey, D., and Edwards, J., editors (1996). *The Essential Distributed Objects Survival Guide.* John Wiley & Sons, New York.