[Mau97] M. L. Mauldin: Lycos: Design Choices in an Internet Search Engine, *IEEE Expert*, January-February 1997. http://www.lycos.com.

[PGW95] Y. Papakonstantinou, H. Garcia Molina, and J. Widom: Object Exchange Across Heterogeneous Information Sources. In *Proceedings of the IEEE International Conference on Data Engineering (ICDE)*, Taipei, Taiwan, March 1995.

[PiS94] T. Pirlein and R. Studer: KARO: An Integrated Environment for Reusing Ontologies. In L. Steels et al. (eds.), *A Future for Knowledge Acquisitio*n, Lecture Notes in Artificial Intelligence (LNAI), no 867, Springer-Verlag, Berlin, 1994.

[SAB+95] V. S. Subrahmanian. S. Adali, A. Brink, J. J. Lu, A. Rajput, T. J. Rogers, R. Ross, C. Ward: *HERMES: A Heterogeneous Reasoning and Mediator System*, Technical Report, University of Maryland, 1995.

[SeE97] E. Selberg and O. Etzioni: The MetaCrawler Architecture for Resource Aggregation on the Web, *IEEE Expert*, January-February 1997. http://www.metacrawler.com.

[SPK+96] B. Swartout, R. Patil, K. Knight, and T. Russ: Toward Distributed Use of Large-Scale Ontologies. In *Proceedings of the 10th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop (KAW-96)*, Banff, Alberta, Canada, November 9-14, 1996.

[Wie92] G Wiederhold: *Mediators in the Architecture of Future Information Systems*, IEEE Computer, 25(3):38-49, 1992.

way enabling its automated manipulation. In the paper, we presented ontologies as means to annotate WWW documents with semantic information. We take the metaphor of a newsgroup (i.e., group ontology) to illustrate the idea of introducing subnets in the WWW. A group-specific ontology is used to annotate web documents giving them a semantics that can be accessed by queries. Different ontologies can be used to annotate different subnets, i.e. to organise the access by semantic queries. Currently, we investigate different systems for integrating heterogeneous information sources and web-querying systems to implement our approach.

Acknowledgements. We thank Richard Benjamins, Stefan Decker, Rainer Perkuhn and two anonymous reviewers for their helpful comments.

6 References

[ACH+93] Y. Arens, C. Y. Chee, C.-N. Hsu and C. Knoblock: Retrieving and Integrating Data From Multiple Information Sources, *International Journal of Intelligent Cooperative Information Systems*, 2(2):127--158, 1993.

[AKS96] Y. Arens, C. Knoblock, and W. Shen: Query Reformulation for Dynamic Information Integration, *Journal of Intelligent Information Systems*, Special Issue on Intelligent Information Integration, 6(2/3), 1996.

[CHS91] C. Collet, M. N. Huhns, and W.-M. Shen: Resource Integration Using a Large Knowledge Base in Carnot, *IEEE Computer*, December 1991.

[DG97] O. M. Duschka and M. R. Genesereth: Infomaster—An Information Integration Tool. In *Proceedings of the International Workshop Intelligent Information Integration during the 21st German Annual Conference on Artificial Intelligence*, Freiburg, Germany, September 9-12, 1997.

[FFR97] A. Farquhar, R. Fikes, and J. Rice: The Ontolingua Server: a Tool for Collaborative Ontology Construction, *International Journal of Human-Computer Studies (IJHCS)*, 46(6):707—728, 1997.

[Gru93] T. R. Gruber: A Translation Approach to Portable Ontology Specifications, *Knowledge Acquisition*, 5(2), 1993.

[GF92] M. R. Genesereth and R. E. Fikes: *Knowledge Interchange Format, Version 3.0, Reference Manual*, Report Logic-92-1, Stanford University, June 1992.

[HSW97] G. van Heijst, A. T. Schreiber, and B. J. Wielinga: Using Explicit Ontologies in Knowledge-Based System Development, *International Journal of Human-Computer Interaction (IJHCI)*, 1997.

[KS95] D. Konopnicki and O. Shmueli: W3QS: A Query System for the WWW. In *Proceedings of the 21th International Conference on Very Large Databases (VLDB-95)*, Zürich, Switzerland, 1995.

[LRO96] A. Y. Levy, A. Rajaraman, and J. J. Ordille: Query-Answering Algorithms for Information Agents. In *Proceedings of the AAAI-96*, Portland, Oregon, 1996.

[LSR96] S. Luke, L. Spector, and D. Rager: Ontology-Based Knowledge Discovery on the World-Wide Web. Proceedings of the *Workshop on Internet-based Information Systems* at the *AAAI-96*, Portland, Oregon, 1996.

[LSR+97] S. Luke, L. Spector, D. Rager, and J. Hendler: Ontology-based Web Agents. In P*roceedings of First International Conference on Autonomous Agents*, 1997.

[LSS96] L. V. S. Lakshmanan, F. Sadri, I. N. Subramanian: A Declarative Language for Querying and Restructuring the Web. In *Proceedings of the 6th International Workshop in Research Issues in Data Engineering—Interoperability of Nontraditional Database System (RIDE-NDS96)*, New Orleans, Louisiana, February 26-27, 1996.

[Mac90] MacGregor: *LOOM Users Manual*, ISI/WP-22, USC/Information Sciences Institute, 1990.

[MMM] A. Mendelzon, G. A. Mihaila, and T. Milo: Querying the WWW, to appear in *Journal of Digital Libraries (JODL).*

? Researcher(X) & firstName(X,“Richard“) & lastName(X,“Benjamins“)
& cooperates(X, Y)

4 Related Work

The idea of using ontologies to annotate information in the WWW is part of the SHOE-approach [LSR96], [LSR+97]. 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 in 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) it can be used for a semantically guided search through the web. We present a much more restricted approach because our approach is suitable for homogeneous intranets and subnets of the WWW created by a community that agree upon a common ontology. As a consequence we can provide the entire ontology used for annotation to the questioner and we can deliver complete answers. Finally, we want to 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. It consists of class hierarchies and relations. No attributes, attribute inheritance, and rules are provided in SHOE.

During the last years, a number of approaches have been developed that support the integration of *distributed* and *heterogeneous* information sources: CARNOT [CHS91], Infomaster [DG97], Information Manifold [LRO96], HERMES [SAB+95], SIMS [ACH+93], and TSIMMIS [PGW95]. Instead of assuming a global data schema such systems have a *mediator* [Wie92] that translates user queries into sub-queries on the different information sources and integrates the sub-answers. Wrappers and content descriptions of information sources provide the connection of an information source to the mediator. Mediators use information source descriptions and general domain information (used to map queries on sources descriptions) that corresponds to group ontologies in our approach. Infomaster uses KIF [GF92], Information Manifold uses CARIN-CLASSIC [LSR+97], and SIMS uses LOOM [Mac90] as languages for defining such mediators. These languages can also be used to define rich ontologies and we will investigate whether one of these approaches can be used to realize our ontologist. We investigate query languages like W3QS [KS95], WebSQL [MMM], and WebLog [LSS96] for the technical aspects of querying the web via the web crawler (i.e., for building the wrapper).

5 Conclusions

Viewing the WWW as a knowledge-based system requires means to represent knowledge in a

3) At his O-page he has to define a sub-index that contains all URLs of web documents containing his ontological annotated knowledge units.
4) He has to actually use the ontology to annotate his bits of knowledge.

When answering a query, the web crawler consults its provider index before visiting the listed O-pages to guide its search in the WWW. Receiving a query from the ontologist it checks each O-page whether it contains a related fragment of the ontology. If it does, each URL of the O-page is investigated. The web crawler then delivers all entities matching the query back to the ontologist. After collecting the factual knowledge from the web the ontologist selects those facts and their consequences that fulfill the entire complex query and sends it back to the client.

Finally we will provide some example queries to illustrate the functionality of the broker. Given the ontology as defined in section 2 we may ask for the research projects of a research group:

? ResearchProject(performedBy:“http://www.aifb.uni-karlsruhe.de/WBS/“)

If we are interested in the publications of one of our colleagues we could ask:

? Researcher(X) & firstName(X,“Richard“) & lastName(X,“Benjamins“) &
Publication(author:Y,title:Z) & X ∈ Y

Here we get all publications in return having a co-author with name Richard Benjamins. Finally, we may ask for all colleagues Richard is cooperating with:

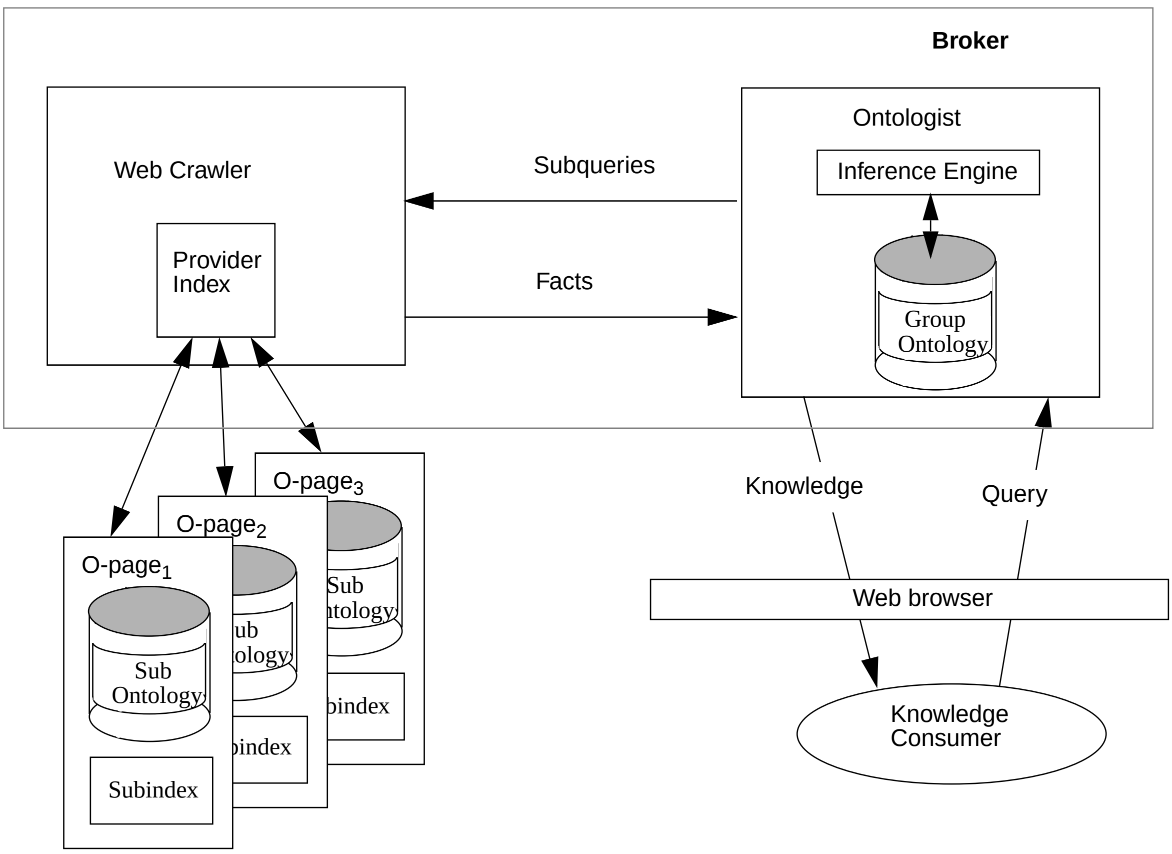
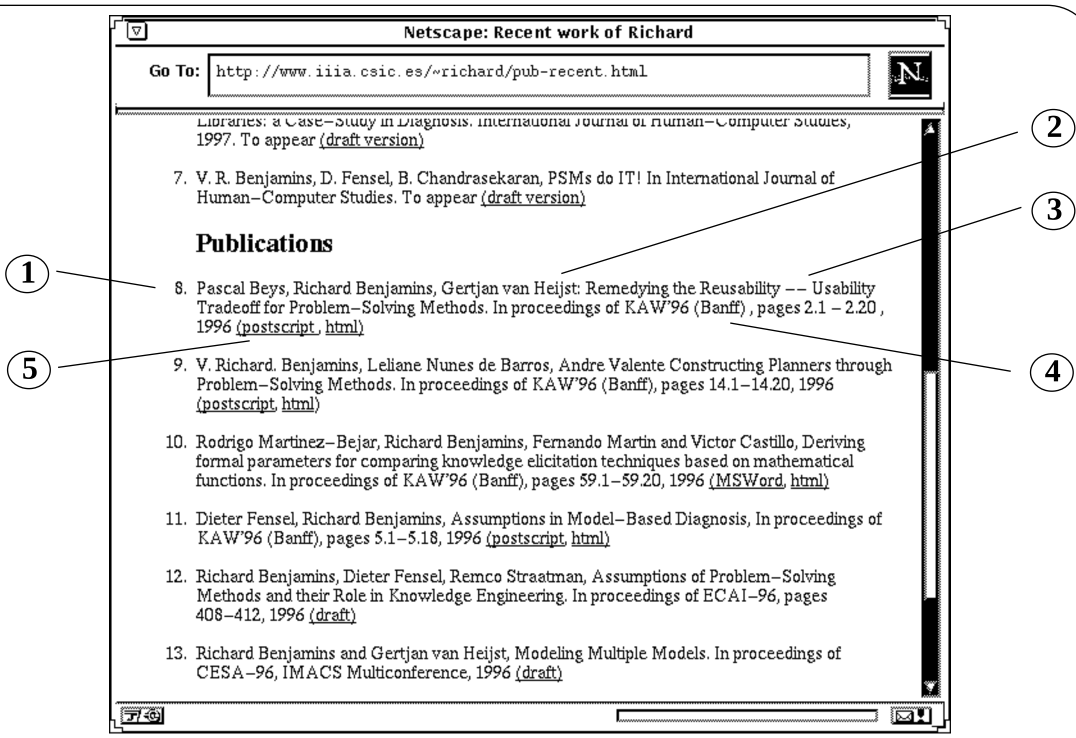


Fig. 3 The architecture of an ontological brokering service for the WWW.



Netscape: Recent work of Richard

Go To: http://www.iiia.csic.es/~richard/pub-recent.html

Libraries: a Case-Study in Diagnosis. International Journal of Human-Computer Studies, 1997. To appear (draft version)

7. V. R. Benjamins, D. Fensel, B. Chandrasekaran, PSMs do IT! In International Journal of Human-Computer Studies. To appear (draft version)

Publications

8. Pascal Beys, Richard Benjamins, Gertjan van Heijst: Remedying the Reusability -- Usability Tradeoff for Problem-Solving Methods. In proceedings of KAW'96 (Banff), pages 2.1 - 2.20, 1996 (postscript, html)
9. V. Richard. Benjamins, Leliane Nunes de Barros, Andre Valente Constructing Planners through Problem-Solving Methods. In proceedings of KAW'96 (Banff), pages 14.1-14.20, 1996 (postscript, html)
10. Rodrigo Martinez-Bejar, Richard Benjamins, Fernando Martin and Victor Castillo, Deriving formal parameters for comparing knowledge elicitation techniques based on mathematical functions. In proceedings of KAW'96 (Banff), pages 59.1-59.20, 1996 (MSWord, html)
11. Dieter Fensel, Richard Benjamins, Assumptions in Model-Based Diagnosis, In proceedings of KAW'96 (Banff), pages 5.1-5.18, 1996 (postscript, html)
12. Richard Benjamins, Dieter Fensel, Remco Straatman, Assumptions of Problem-Solving Methods and their Role in Knowledge Engineering. In proceedings of ECAI-96, pages 408-412, 1996 (draft)
13. Richard Benjamins and Gertjan van Heijst, Modeling Multiple Models. In proceedings of CESA-96, IMACS Multiconference, 1996 (draft)

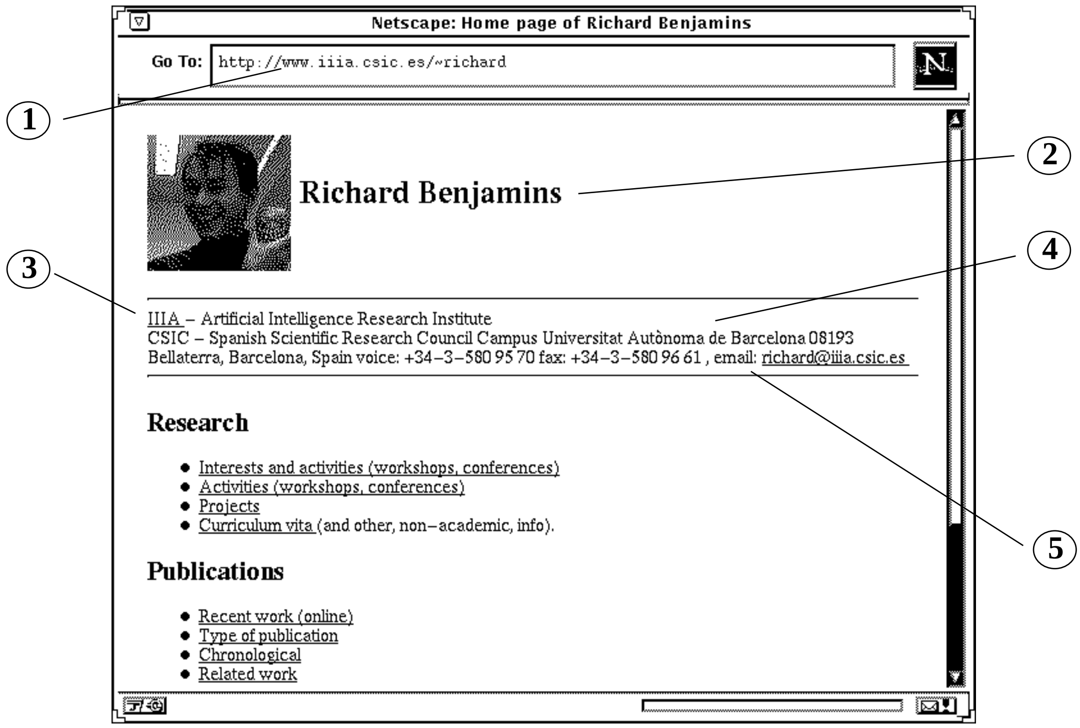

```
<html>
<head><title>Recent work of Richard</title></head>
<body>
   ...
<A name=publications> <h2>Publications </h2></A>
<li>
<A name="KAW96"
   onto="CLASS WorkshopPaper(1)(      author=http://www.swi.psy.uva.nl/usr/beys/home.html(2)
                                      author=http://www.iiia.csic.es/~richard
                                      author=http://www.swi.psy.uva.nl/usr/gertjan/home.html">
   Pascal Beys, Richard Benjamins, Gertjan van Heijst:
</A>
<A onto="ATTRIBUTE #KAW96 (title)">
   Remedying the Reusability -- Usability Tradeoff for Problem-Solving Methods.(3)
</A>
In <A onto="ATTRIBUTE #KAW96 (procedingsTitle)"> proceedings of KAW'96 (Banff) </A>(4),
<A onto="ATTRIBUTE #KAW96(conference=http://www.ksi.cpsc.ucalgary.ca/KAW/KAW.html)"></A>
pages
   <A onto="ATTRIBUTE #KAW96 (firstPage)"> 2.1 </A>-
   <A onto="ATTRIBUTE #KAW96 (lastPage)">  2.20 </A>,
<A onto="ATTRIBUTE #KAW96 (year)"> 1996 </A>
<A HREF="../postscripts/banff96-reuse.ps"
   onto="RELATION #KAW96 (onlineVersion)"> (5)
   (postscript
</A>,
<A HREF="http://ksi.cpsc.ucalgary.ca/KAW/KAW96/beys/setup.html"
   onto="RELATION #KAW96 (onlineVersion)"> (5)
   html)
</A>
```

Fig. 2 The publication page of Benjamin Richards.

makes use of the ontology underlying the ontologist. For this it uses a provider index that contains an URL for each knowledge provider who makes use of the ontology to annotate his knowledge units.

Each provider of an ontological annotated knowledge chunk has to do four things:

1) He has to define one O-page and has to register its URL in the provider index of the web crawler of the ontology.
2) At his O-page he has to define the subset of the group ontology that he will use to annotate his knowledge chunks.




```
<html><head><TITLE> Home page of Richard Benjamins </TITLE></head>
<body>
<H1> <A onto="CLASS Researcher"(1)></A> <IMG ALIGN=CENTER SRC="rich1.gif">
        <A onto="ATTRIBUTE http://www.iiia.csic.es/~richard (firstName)"> Richard </A>
        <A onto="ATTRIBUTE http://www.iiia.csic.es/~richard (lastName)"> Benjamins </A>(2)
</H1>
<p><HR>
<A HREF="../../IIIA.html" onto="ATTRIBUTE http://www.iiia.csic.es/~richard (affiliation)"> IIIA </A>(3)
- Artificial Intelligence Research Institute<br>
CSIC - Spanish Scientific Research Council
<A onto="ATTRIBUTE http://www.iiia.csic.es/~richard (address)">(4)
        Campus Universitat Aut&ograve;noma de Barcelona 08193 Bellaterra, Barcelona, Spain
        voice: +34-3-580 95 70 fax: +34-3-580 96 61
</A>, email: (5)
<A HREF="mailto:richard@iiia.csic.es" onto="ATTRIBUTE http://www.iiia.csic.es/~richard (email)">
richard@iiia.csic.es </A>
<HR>
...
<h2> Publications </h2>
<ul>
<LI> <A HREF="pub-recent.html"> Recent work (online)</A>
...
```

Fig. 1 The homepage of Richard Benjamins.

Using this variant of defining attributes requires the definition of exactly one attribute per anchor. The attributes defined in this way have to be atomic, i.e. must be strings or numbers etc. If the attribute is not atomic one can omit the attribute value as well but has to provide an href-part in the anchor definition. This link then defines the missing attribute value, e.g.

defines the entity identified by "http://www.iiia.csic.es/~richard" as the author of a publication.

The same technique can be applied to relations. To avoid multiple input of identical texts the last argument of a relationship definition can be omitted in some cases. If this entity is of atomic type the text between <a> and becomes the last argument; if the last argument is defined as a concept and a href-part is given when defining the anchor, the href-URL is interpreted as the missing entity.

2.3 Example of an Annotated Web Page

In this section we will present two HTML pages annotated with ontological information. The first presented page is the home page (see Figure 1) of Richard Benjamins, a researcher at the Artificial Intelligence Research Institute (IIIA from the Spanish Scientific Research Councel. His home page contains information defining him as a researcher (1), his first and last name (2), his affiliation (3), and his address(4) and e-mail (5). He is identified by the URL of his home page (http://www.iiia.csic.es/~richard). All attribute and relationship definitions contain this URL as key argument.

On this homepage a link pointing to Richards' publications is defined, which serves as the second example. This page contains a long list of publications (see Figure 2). These publications are identified by unique keys starting with http://www.iiia.csic.es/~richard/pub-recent.html followed by a number sign and their named, as defined within an anchor tag. Within the page they are simply referred to by a number sign, followed by their name, e.g. #KAW96. The example in Figure 2 displays the ontological annotations of one article which appeared in the KAW proceedings 1996. The publication is classified as an WorkshopArticle (1) which was published 1996. Mr. Benjamins and two of his colleagues are defined as the authors (2), the title of the article is defined (3), the proceedings title and the pages in the proceedings are stated (4), and the article is related to on-line versions which are available via www as postscript or as HTML documents (5).

3 The Ontology-based Broker

The general architecture of an ontology-based brokering service for the WWW is shown in Figure 3. The broker has to communicate with clients that have a query asking for some knowledge using web browsers like Netscape and Explorer. The core of the broker service consists of two main elements: an *ontologist* and a *web crawler*.[2] The *ontologist* receives the query and translates it into subqueries for facts that are passed to the web crawler. In return it receives a set of ground facts. The web crawler searches through a fragment of the WWW that

2. For efficiency reasons a third element is necessary that minimises the effort of query execution. However, we will not discuss this query planner here and refer to [AKS96], [LRO96].

To define the fact that an entity belongs to a certain concept of the ontology, one uses the HTML-tag

<a ... onto=“CLASS *class-name*“> ...

E.g. on the home page of Mr. Richard Benjamins (cf. example below) he could define

 ...

If the anchor tag does not contain a name attribute the whole page is defined as belonging to the given class; the page‘s URL then becomes the entity‘s unique identifier. If a name-attribute is given, e.g.

 ...

the concept is defined for exactly this anchor; its unique identifier is then composed (according to URL syntax) of the URL of the web-page plus the number-sign (#) plus the name of the anchor itself, e.g. http://www.iiia.csic.es/~richard/pub-recent.html#KAW96.

2.2.2 Reference of an ontological annotation

Besides scope, must must define the ontological entity that gets characterized. When defining a concept, one can define its attributes within the same tag, as in

<a ... onto=“CLASS *class-name*(*attribute*$_1$*=value*$_1$*, ... attribute*$_n$*=value*$_n$)“> ...

or the attributes can be defined within individual anchor-tags. In the latter case the referred-to entity must be given in terms of its identifier.

<a ... onto=“ATTRIBUTE *entity-id(attribute*$_1$*=value*$_1$*, ... attribute*$_n$*=value*$_n$)“> ...

Relations are defined in roughly the same way as attributes. To define a relation between *n* entities the following definition suffices.

<a ... onto=“RELATION *relation_name*(*entity*$_1$*, entity*$_2$*, ... entity*$_n$)“> ...

The entities contained in this definition must be given in terms of their identifiers, e.g.

<a onto=“RELATION *cooperates*(http://www.iiia.csic.es/~richard,
http://www.aifb.uni-karlsruhe.de/~dfe)“> ...

This ontological statement relates the two entities http://www.iiia.csic.es/~richard and http://www.aifb.uni-karlsruhe.de/~dfe according to the cooperation-relationship. These entities represent two persons, identified by their respective home pages.

2.2.3 Value of an ontological annotation

Finally, one has to provide means to characterize ontological entities. To avoid redundancy by writing information twice on a web-page (first as ontological information, and second as human-readable and browser-presentable text) the definition of attributes can be abbreviated. Instead of giving the attribute values within the anchor <a> it can be included between the <a> and part of an anchor, e.g.

<a ... onto=“ATTRIBUTE *entity-id(attribute*)“> *value*

```
Project(projectName:STRING, head:Person)
    ResearchProject(performedBy:ResearchGroup)
    DevelopmentProject
    SoftwareProject
Event(  eventTitle:STRING, location:STRING, date:DATE,
        programCommittee: SET OF Person, orgCommittee: SET OF Person))
    Conference(series:STRING, number:INTEGER)
        SEKE(series="SEKE",
            eventTitle="Intl. Conference on Software and Knowledge Engineering")
    Workshop(series:STRING, number:INTEGER)
        KAW(series="KAW", location="Banff, Canada",
            eventTitle="Workshop on Knowledge Acquisition, Modeling,
                    and Management"
        EKAW(series="EKAW",
            eventTitle="European Workshop on Knowledge Acquisition, Modeling,
                    and Management"
        KEML(series="KEML",
            eventTitle="Knowledge Engineering: Methods & Languages"
```

After defining the concept-taxonomy some **relationships** between concepts follow.

```
cooperates(Researcher, Researcher)
isMemberOf(Researcher, ResearchGroup)
onlineVersion(Publication, OnLinePublication)
hasPublications(Person, Publication)
hasProjectInfo(Project, Publication)
participatesIn(Person, Event)
```

Rules - as parts of the ontology - enable an ontology-based search-engine to infer knowledge implied by the facts stored in the web.The following example defines the symmetry of cooperation:

```
cooperates(X, Y) ←cooperates(Y, X)
```

2.2 Annotating Web-Pages with Ontological Information

Because we want to annotate HTML-Pages with semantic information as structured in our ontology, we have to define a means to express certain facts within web pages to enable automatic inference. These facts are coded within HTML-tags, esp. the anchor tag. The anchor tags (i.e. <a ...> ...) are naturally used to define hypertext links or to define named locations as targets for hypertext links. To connect ontological information to parts of HTML-pages, we add another attribute to the anchor tag, namely the *onto*-attribute. Any information given in this onto-element annotates the whole anchor.

2.2.1 Scope of an ontological annotation

All ontological concepts and relationships have to refer to identifiable entities on the web. The easiest way to identify an entity on the web is its URL (Uniform Resource Locator). This URL represents—by definition—a unique way to determine which entities are referred to by concepts or relationships.

2.1 An Example Ontology

This section defines an ontology which contains concepts and relationships capable of describing the KA research community. The ontology contains many concepts that are quite general, esp. the subhierarchies Person and Publication. Therefore, they should be reusable for other *islands in the WWW*. The definition of the ontology for the KA-research community is provided below. Firstly **ontological concepts** are defined using an *is-a*- taxonomy, and appropriate attributes are specified.[1]

```
Organization(name:STRING, location:STRING)
    University
    Department(partOf:University)
    Institute(partOf:Department)
    ResearchGroup(partOF:Institute, member:SET OF Person)
    Enterprise
Person(firstName:STRING, lastName:STRING, address:STRING, email:STRING)
    Employee(affiliation:Organization)
        AcademicStaff
            Researcher
            Lecturer
        AdministrativeStaff
            Secretary(secretaryOf:Employee)
            TechnicalStaff
    Student(studiesAt:University)
        DoctoralStudent(supervisor:AcademicStaff)
Publication(    author:SET OF Person, title:STRING, year:NUMBER)
    Book(publisher:STRING, editor:SET OF Person)
    Article
        TechnicalReport(series:STRING, number:NUMBER, organization:Organization)
        JournalArticle(journal:Journal, firstPage:NUMBER, lastPage:NUMBER)
        ArticleInBook(book:Book, firstPage:NUMBER, lastPage:NUMBER)
        ConferencePaper(conference:Conference, proceedingsTitle:STRING,
                        firstPage:NUMBER, lastPage:NUMBER)
        WorkshopPaper( workshop:Workshop, proceedingsTitle:STRING,
                        firstPage:NUMBER, lastPage:NUMBER)
    Journal(editor:Person, volume:NUMBER, number:NUMBER)
        IJHCS( title="International Journal of Human Computer Studies")
        DKE(title="Data and Knowledge Engineering")
        IEEE_Expert(title="IEEE Expert")
    OnLinePublication
```

The subconcepts of Journal, i.e. IJHCS, DKE, and IEEE_Expert contain each issue of the corresponding journal as an instance.

---

1. To increase the readability of the presented ontology we will first define some notational conventions: The ontology is given as an indented list of concepts. The indentation indicates *is-a* relationships between concepts, e.g. a Employee *is-a* Person etc. Words starting with an uppercase letter represent **concepts** of the ontology, e.g. Organization. All-uppercase words are **atomic types**, i.e. STRING, NUMBER, DATE, BOOLEAN. The lists appearing behind certain concepts contain **attribute definitions** for these concepts following the format <attributeName>:<attributeType>. Attribute types are either concepts (e.g. Employee(affiliation:Organization)) or atomic types (e.g. Person(lastName:STRING)). All-lowercase words represent **attribute names** or names of **relationships** in general. All attributes and relationships defined for a (super)concept are inherited transitively to all of its descendants (subconcepts) of the *is-a* hierarchy.

Deriving semantic information from sentences in natural language automatically manner is still an unsolved problem. Inference by keyword search may deliver some results but it also results in a lot of unrelated information and at the same time it may miss a lot of important information. As [LSR96], [LSR+97] point out searching for a person with name *Cook* leads to thousands of pages about cooking and searching for cooking advice forces the search machines to follow links from the home page of Mr. and Mrs. Cook even if they have never put any knowledge about cooking on the web. [LSR96] and [LSR+97] propose *ontologies* to improve the automatic inference support of the knowledge base WWW. An ontology provides "an explicit specification of a conceptualization" [Gru93]. Ontologies are discussed in the literature as means to support knowledge sharing and reuse ([PiS94], [FFR97], [HSW97]). This approach to reuse is based on the assumption that if a modelling schema - i.e. an ontology - is explicitly specified and agreed upon by a number of agents, it is then possible for them to share and reuse knowledge. Standardizing the syntactical way in which semantic information is presented allows the automatic derivation of semantic information via syntactical manipulation.

It is clear that such an ontology will only be shared by a small group of web users. We cannot expect that ontologies will be used by everybody and even if everybody would use ontologies to annotate his web pages it will be hardly ever possible to negotiate on a worldwide-used standard for representing knowledge about all possible subjects. Therefore, we use the *metaphor of a newsgroup* to define the role of such an ontology. It is used by a group of people that share a common subject and a related point of view on this subject. Thus it allows them to annotate their knowledge to enable automatic inference based on the shared ontology. So different group ontologies define different *islands of meaning* in the entire WWW ocean. Such an island defines a knowledge-based subsystem of the web that can be supplemented with an inference engine based on its ontology. In the long term, if several of these islands arise one can wonder whether they can be put under a common umbrella of more abstract ontologies that contains the different ontologies as specializations [SPK+96].

The contents of the paper is organised as follows. We will present an ontology as a means to enrich HTML documents for presenting semantic information about these documents in section 2. We will use the representation of a research community in the WWW for illustrating our ideas. HTML documents become annotated by adding additional attributes to the anchor tag of HTML. In section 3, we sketch the general architecture for an ontology-based broker for the WWW. We discuss two main parts: an ontological reasoner that executes the inferences and a web crawler that collects input from the web. Related work is discussed in section 4 and we provide conclusions in section 5.

2 Ontological Annotations for Web Documents

In this section we will (1) define an ontology to represent one potential group of users, the research community of KA (Knowledge Acquisition) in the WWW, (2) describe how to annotate web-pages using this ontology, and (3) give an illustration of an annotated web-page. Our design rationale has been to ensure that the annotation can be smoothly integrated into HTML documents. That is, there is no requirement to completely annotate WWW-documents and we make use of existing textual information as much as possible.

In *Proceedings of the International Workshop Intelligent Information Integration during the 21st German Annual Conference on Artificial Intelligence*, Freiburg, Germany, September 9-12, 1997.

Ontology Groups: Semantically Enriched Subnets of the WWW

Dieter Fensel, Michael Erdmann, and Rudi Studer
University of Karlsruhe, Institute AIFB, 76128 Karlsruhe, Germany
E-mail: {fensel, erdmann, studer}@aifb.uni-karlsruhe.de

Abstract. The World Wide Web (WWW) can be viewed as the largest knowledge-base that has ever existed. However, its support in automated inference is very limited. We present ontologies as means to enrich web documents for representing semantic information for overcoming this bottleneck. Ontologies enable informed search for information as well as the derivation of additional knowledge that is not directly represented in the WWW. Such an ontology can be used by a subgroup of web users that share a common point of interest like e.g. newsgroups in the internet. Therefore, we propose ontologies to be used to annotate intranets or certain subnets within the entire WWW. The paper presents such an ontology that can be used to annotate web documents of a research community. In fact, we have chosen the knowledge acquisition community. Further, an architecture of an ontology-based broker is sketched that can make use of these ontologies for the automatic derivation of knowledge.

1 Introduction

The World Wide Web (WWW) can be seen as the largest knowledge-based system that has ever existed. It contains huge amounts of knowledge about any subject one can think of. HTML documents enriched by multi-media provide knowledge in different representations (i.e., text, graphics, moving pictures, video, sound, virtual reality, etc.). Hypertext links between web documents represent relationships between different knowledge entities. Based on the HTML standard, browsers are available that present the material to humans. Browsers can also use the HTML-links to browse through distributed information and knowledge units. However, taking *the metaphor of a knowledge-based system* as a way to look at the WWW brings the big bottleneck of the web into mind. Its support in automated inference is very limited. Deriving new knowledge from existing knowledge is hardly supported. Actually, the main inference services the web provides are keyword-based search facilities realized by different search engines, web crawlers, web indices, man-made web catalogues etc. (see [Mau97], [SeE97]). Given a keyword, such an engine collects a set of knowledge chunks from the web that use this keyword. This limited inference access to existing knowledge stems from the fact that there exist only two main types of standardization for knowledge representation in the web:

- the HTML standard is used to present knowledge in a (browser and) human-readable way and to define links between different knowledge units.
- mainly the English language is used to represent the knowledge units.