

Conceptual Linking: Ontology-based Open Hypermedia

Leslie Carr, Wendy Hall
Intelligence, Agents, Multimedia
Department of Electronics and Computer Science,
University of Southampton,
Southampton SO17 1BJ, UK
{lac,wh}@ecs.soton.ac.uk

Sean Bechhofer, Carole Goble
Information Management Group
Department of Computer Science
University of Manchester
Oxford Road
Manchester M13 9PL, UK
{seanb, carole}@cs.man.ac.uk

Abstract

This paper describes the attempts of the COHSE project to define and deploy a Conceptual Open Hypermedia Service. Consisting of

- *an ontological reasoning service which is used to represent a sophisticated conceptual model of document terms and their relationships;*
- *a Web-based open hypermedia link service that can offer a range of different link-providing facilities in a scalable and non-intrusive fashion;*

and integrated to form a *conceptual hypermedia* system to enable documents to be linked via metadata describing their contents and hence to improve the consistency and breadth of linking of WWW documents at retrieval time (as readers browse the documents) and authoring time (as authors create the documents).

Keywords

Open hypermedia, link service, ontology, navigation, metadata.

Introduction: concepts and metadata

Metadata is data that describes other data to enhance its usefulness. The library catalogue or database schema are canonical examples. For our purposes, metadata falls into three broad categories:

- *Catalogue information: e.g. the artist or author, the title, a picture's dimensions, a document's revision history;*
- *Structural content: e.g. headings, titles, links; for a picture its shapes, colors and textures;*
- *Semantic content: what the document/picture is about e.g. football, sport, person holding trophy, hope, joy.*

Metadata activities have been a major focus of interest for the WWW community, especially for information providers, publishers and digital libraries. The takeup of the eXtensible Markup Language (XML) has been particularly concerned with its applications for expressing data about documents, and has most recently been used to define the Resource Description Framework (RDF) [26]. The aim of RDF is to provide a standard framework for expressing statements about data objects, especially statements giving information about authors, publishers, version and keyword information (these attributes being standardized as the Dublin Core [31]).

An ontology is a formal model of the kinds of concepts and objects that appear in the real world, together with the relationships between them. Ontologies take a variety of forms, from hierarchical classification schemes such as Yahoo! directories to logic-based models. All these forms include at least a vocabulary of terms and some specification of the meaning of those terms.

Using Metadata for Linking

Providing conceptual content-based information as the attributes of web pages is an important activity, enabling search engines to provide query results that are more pertinent. Currently such concepts are usually simple keywords. Hypermedia systems such as the Distributed Link Service [9, 10] may make use of this information to provide a rudimentary "conceptual hypermedia" by clustering documents with the same tag value keyword for retrieval purposes and linking documents with the same tag value for navigation. Keywords effectively classify documents into clusters that share the same set of keywords, or variations of them if stemming is used.

To achieve the kind of diversity of association required for non-trivial Web applications, documents need to be linked in many dimensions based on their content. Constructing such links manually is inconsistent and error-prone [17]. Furthermore, it obfuscates one of the chief reasons for associating documents; that their contents are similar in some way. Conceptual Hypermedia Systems (CHS) specify the hypertext structure and behaviour in terms of a well-defined conceptual schema [7, 28, 33]. This types documents and links, and includes a conceptual domain model used to describe document content. Consequently, information about the hypertext is represented explicitly as metadata that can be reasoned over, for example, using the domain model as a classification structure to classify the documents; documents that share metadata are deemed to be similar in some way. Authoring links between documents becomes an activity of authoring with concepts; concepts are linked and hence their associated documents are linked.

Open Hypermedia Systems and Link Services

Common usage of the Web involves embedding links within documents in the HTML format; in this sense the Web can be considered a 'closed' hypermedia system. However, there is nothing inherent in the Web infrastructure that prevents hypertext links from being abstracted away from the documents and managed separately, for example, by using XLink's third-party links [14]. In Open Hypermedia Systems (OHS) links are first class objects, stored and managed separately from multimedia data; like documents they can be stored, transported, cached and searched, and their use can be instrumented. OHS have been well researched by the hypermedia community [21, 29] and increasingly Web publishing applications adopt the open hypermedia approach [27, 32].

The DLS provides a powerful framework to aid navigation and authoring and addresses some of the issues of distributed information management [16]. Using an intermediary model [2], the DLS adds links and annotations into documents as they are delivered through a proxy from the original WWW server to the ultimate client browser. It uses a number of software modules to recognize different opportunities for adding various kinds of links to the documents, creating a user-specific navigational overlay that can be used to superimpose a coherent interface to sets of unlinked or insular resources (such as the Eprint archives addressed by the Open Citation project [24]).

The DLS treats link creation and resolution as a service that may be provided by a number of link resolution engines. For example, it uses resolvers which recognize keywords, names of people and bibliographic citations as potential link anchors according to different heuristics and knowledge bases. These link resolvers are hardwired into the monolithic system or chained sequentially [15] so that each one sees the document with links added by the previous resolver. This inherently synchronous arrangement means that any delay is a delay in the critical path of document delivery, hence all processing must be relatively light-weight and tightly coupled.

By contrast, a COHSE needs a Distributed Link Resolution Service (DLRS) to allow link resolvers to be distributed across multiple servers and decoupled from the delivery of the document. The aim is to allow complex computation, such as involved with implementing conceptual inferencing logic to provide added value for document authoring and browsing without impeding the delivery of the core document itself.

COHSE

A terminology-based query system can be added to the portfolio of link resolvers to provide consistent navigation links based on the concepts contained in the contents or meta-data of the multimedia pages that are being browsed.

For the purposes of this demonstration, the link service is integrated into the browser client, and consists of a Java applet that monitors the user's interaction with the browser together with a set of JavaScript functions that manipulate the HTML DOM. The components of the link service (Figure 1) are brought to bear on the web page as soon as it has been received by the browser, and so no longer form an obstacle to the delivery of the document. Once the set of links has been chosen, the page is refreshed and redisplayed.

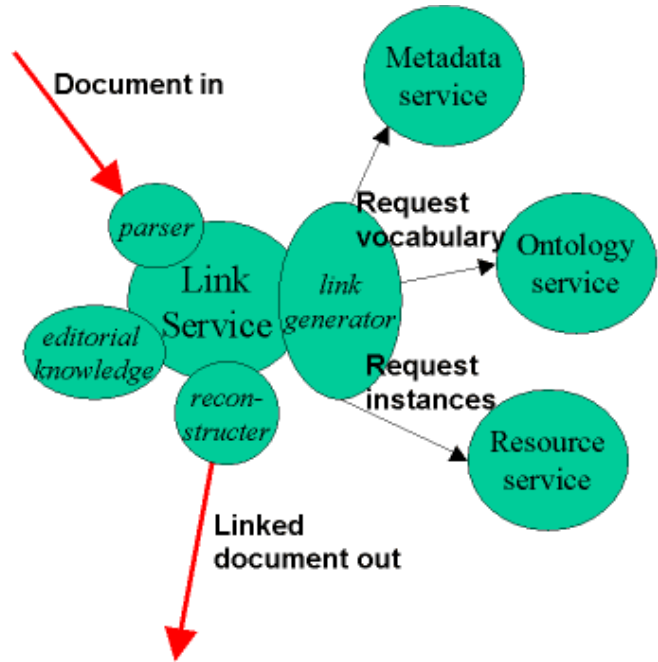


Figure 1: The Ontology Service maps between natural language terms and a concept graph. The Resource Service obtains Web pages representing the concepts. The Link Generator uses the ontology terms to make links. Editorial knowledge is used to prune or expand the links using ontology semantics.

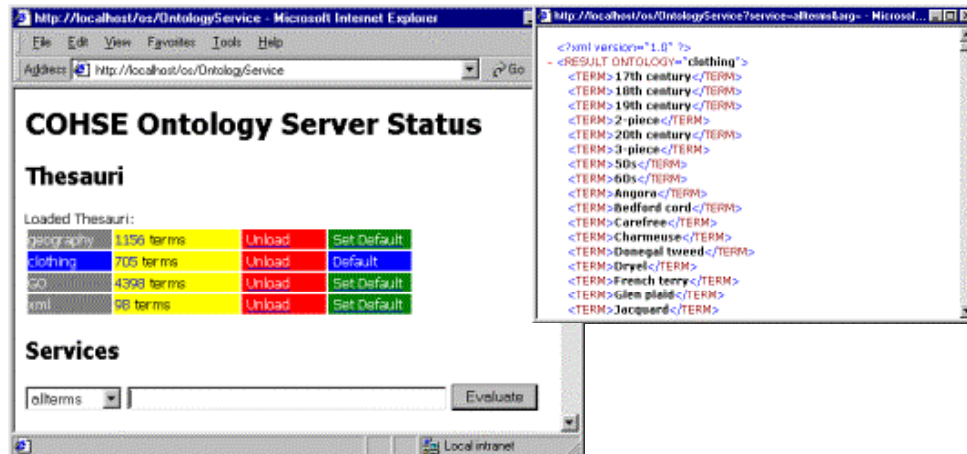


Figure 2: The *Ontology Service* can be queried for terms and concept relationships.

The ontology service (a Java servlet, shown operating in figure 2) manages ontologies, that is to say, sets of concepts that are related together according to some schema. The ontologies currently used take the form of a thesaurus, *i.e.* concepts related by *broader-term*, *narrower-term* and *related-term* relations, and are stored as XML data, with all queries and results being mediated through a simple XML document type.

The link generator module of the link service contacts the ontology service to obtain a complete listing of all the language terms that are used to represent the concepts in the ontology. For each of those terms that are recognised as occurring in the document, the generator first asks the ontology service for a preferred term, and then asks for the preferred term to be mapped onto a concept. Having identified a concept from the strings in the document, the link generator contacts the resource service to obtain a list of documents that contain instances of this concept. At this point, a number of destinations have been identified for a particular link anchor and the editorial module evaluates the number and quality of potential links obtained

from the generator. If the number of links is not consistent with the formation of a *well-linked document*, it will choose to request broader or narrower terms from the ontology service in order to expand or cull the set of anchor destinations. When all the terms in the whole document have been processed, the constructor can add hypertext links with particular presentation styles and behaviours. Figure 3 shows how links are added to an example document (the COHSE control panel appears in figure 3b and the link behaviour in 3c and 3d is shown in debugging mode, so that the link expands within the document text).

The metadata service is another independent servlet which allows documents to be decorated with metadata: language terms from a specific ontology. The service can either harvest specific tags from the documents themselves or apply external 'metadata links' to a read-only document from an independent linkbase. The effect is to declare that a whole document, or any range within it, should be processed with a specific ontology, or that a particular region in the content corresponds to a particular term in the ontology.

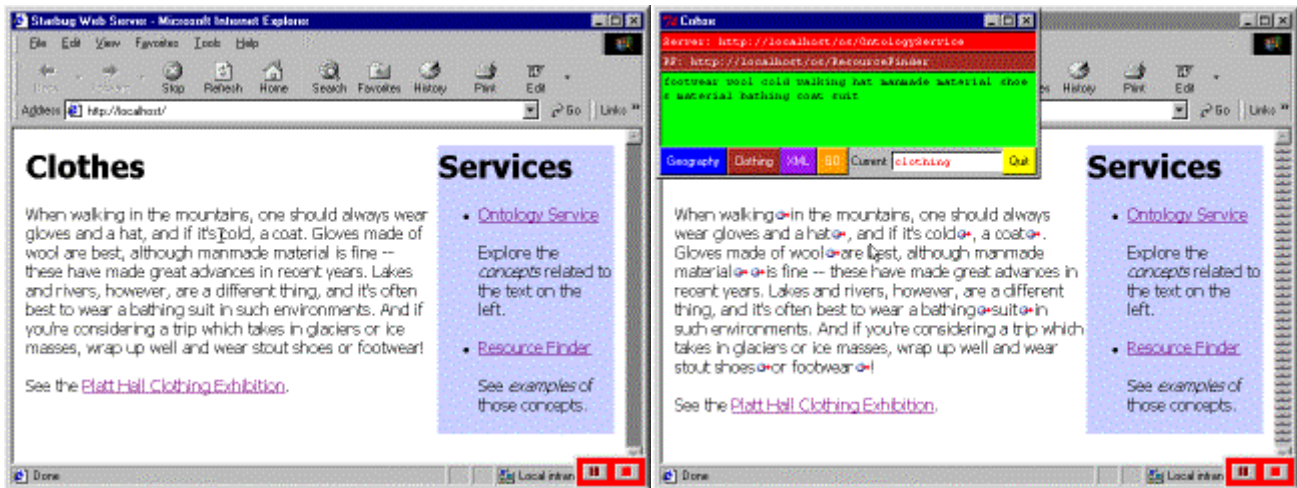


Figure 3a: A page about clothes...

Figure 3b: ... is linked against the clothing ontology...

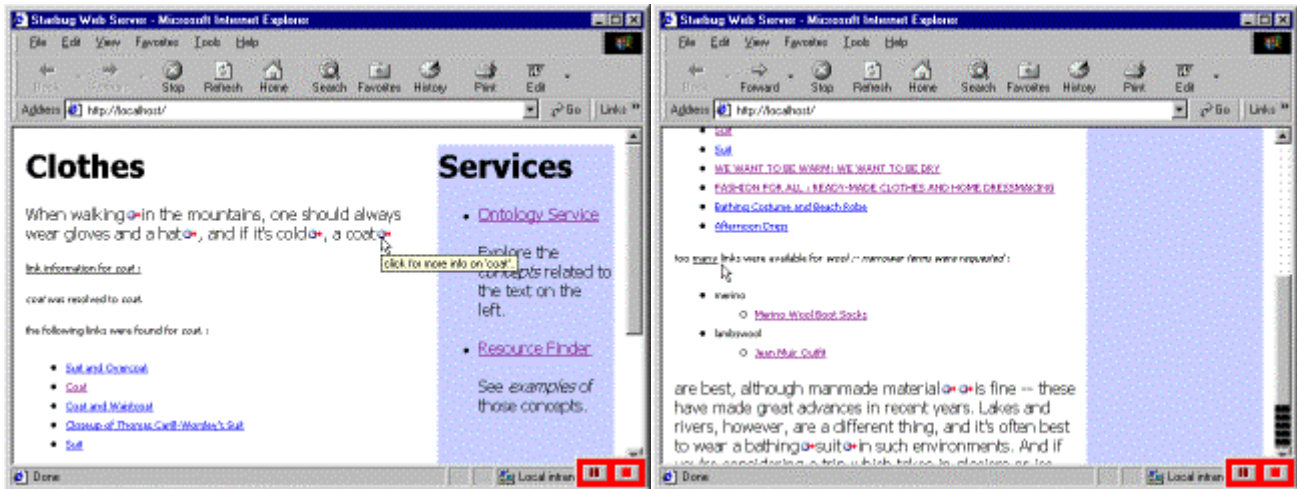


Figure 3c: ... some terms link well with the correct number of destination links...

Figure 3d: ... while some are pruned.

The novel part of the link resolution process is the use of the editorial knowledge component to take advantage of the implicit structure of the ontology to make informed decisions about the kind of links to choose. By making a selection from a set of 'narrower terms' the list of links can be usefully reduced whilst broadening the recognised concept can be used as a strategy to increase the number of links.

Alternative Approaches to Concepts

The COHSE uses a predefined ontology to choose candidate anchors for creating links. This section lists some of the other systems that concern themselves with manipulating concepts, the different ways of representing them and the various modes for deploying them.

Meta Tags

HTML's <META> tags allow authors to specify information about web resources. This is highly uncontrolled though, as the tags contain unconstrained terms and are used for a variety of purposes (indicating author, how the page was generated, content, special things for particular applications and so on). In this case, the metadata is tightly bound to the documents. In order to discover the metadata we must examine the document itself. There is no central metadata repository, but process such as web robots can be sent off to harvest and cache the metadata.

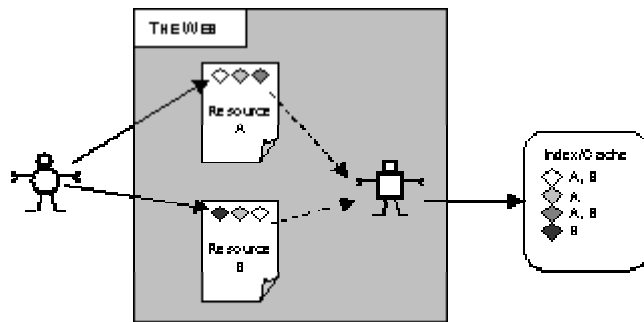


Figure 4: Simple Meta Tags

The index duplicates the metadata from the resources, but provides easy access without having to go out to the Web. Of course with such an approach, the issue of maintenance is crucial. It is difficult to tell whether the index is up to date. The repository or index is also very simple — in general there's no ontological structure, just a list of arbitrary keywords.

Yellow Pages

With a Yellow Pages service, such as Yahoo!, pages are classified according to their content. A taxonomy or hierarchy is normally used, with subject areas being broken down. This is generally achieved by hand, with both the classification hierarchy and the categorization of the pages done manually. The pages themselves are unaltered, so the situation here is that the metadata is stored externally from the documents and has no real link to the documents, other than through the classification.

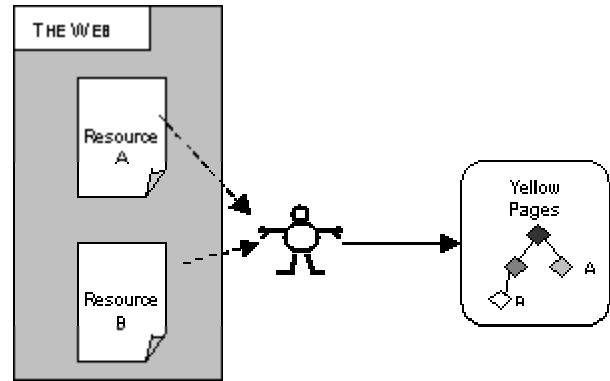


Figure 5: Yellow Pages

Again, this provides a "snapshot" of the situation, so there may be problems that the classification is not up to date. There's also very little automation going on here — this approach may be geared towards supporting humans trying to locate resources rather than providing machine-readable knowledge.

SHOE

The Simple HTML Ontology Extension (SHOE) [22, 23] has been developed by the Parallel Understanding Systems Group in the Department of Computer Science at the University of Maryland. SHOE provides mechanisms that allow the definition of ontologies and the assertion of claims about resources with respect to those ontologies.

Assertions about particular web pages (or resources) are included within pages as mark-up using an HTML based syntax, with a META tag used to inform any agents that the page uses SHOE. The assertions take the form of instance descriptions, asserting membership of classes and relationships between the instances.

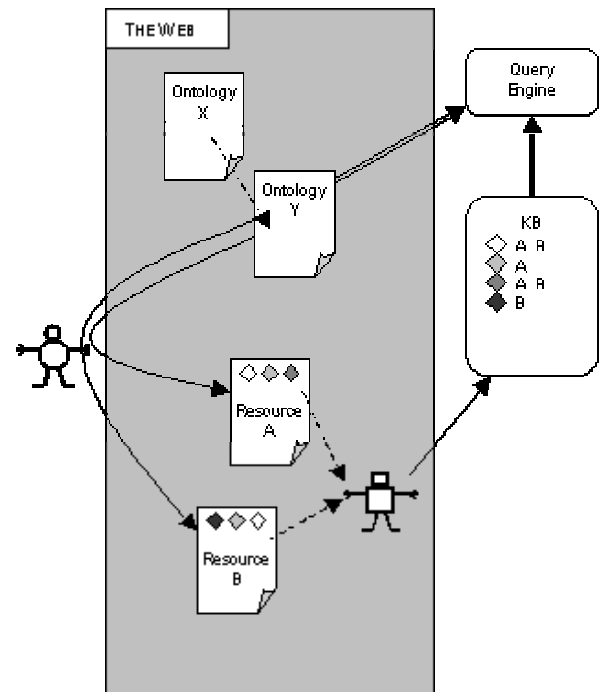


Figure 6: SHOE

In the SHOE model, as shown in figure 6, information is spread around. The metadata is attached explicitly to the documents (through the use of the SHOE HTML extension), but can then be gathered in one place by a robot for later query. Presumably, if one had a SHOE-enabled browser the user could also examine the metadata in situ once they had reached a SHOE annotated page.

Ontobroker

Ontobroker (or On2broker as it is also now known) [13, 19], is a system and architecture from AIFB, Karlsruhe. It is similar in many ways to SHOE and allows the annotation of web pages with ontological metadata. It provides a more expressive framework for the ontologies, using Frame-Logic for the specification of ontologies, annotation data and queries. Ontobroker and SHOE share some characteristics. They both use annotation of the documents themselves, and then rely on web crawlers (crawling through a well-defined docuverse) to harvest the metadata, storing it in a knowledge base. The KB is then queried using the ontology as a schema for query forming. They do differ in a couple of aspects. SHOE provides ontology extension mechanisms and explicitly places the ontologies on the Web. It's less clear how one gains access to the Ontobroker ontologies or how one makes the link between the instance markup and the ontology it applies to.

Karina

In Karina [12], an ontology is used to describe the content of documents in a multimedia repository. This metadata is then used to construct or author a presentation that fits the needs of a particular user. Karina is using the metadata as an index (resource discovery). In Karina, however, the emphasis is that the ontology will then be used in order to structure the results. Karina is thus closely related to COHSE, in that the ontology is being used to produce a structure.

RDF

RDF (Resource Description Framework) [26] differs from systems like SHOE and Ontobroker as it's a framework rather than a particular implemented system — systems such as Ontobroker may use RDF as a representation format. It's useful to compare it here though. RDF provides a framework which allows us to talk about metadata. The RDF data model is based around ideas of triples, i.e. an object, relationship and value. The intention with RDF is that metadata can be held separately from the documents (using the "about" attribute of RDF), separating

the metadata from the document. With RDF, though, the RDF documents themselves are on the Web, so the repository is accessible.

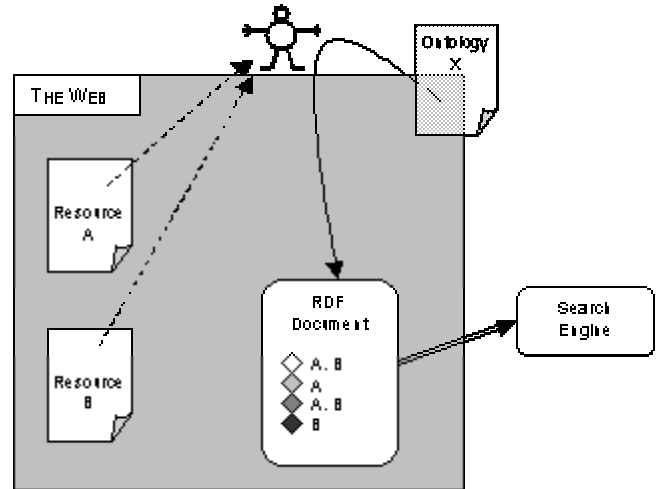


Figure 7: RDF

Figure 7 shows how RDF descriptions might work. Note again that as RDF is simply a framework, this is only one of a number of possible ways that things could be put together. RDF is less prescriptive in its ontology specification — indeed most uses of RDF so far seem to be in order to specify minimum data sets such as the Dublin Core [31], so it is less clear here whether the ontology might sit on the web (as with, say SHOE), or somewhere else. We could consider the RDF document as forming a “knowledge base” or repository with collected metadata about a number of resources.

COHSE

COHSE combines the Distributed Link Service (DLS) [7] architecture with a conceptual model to provide Conceptual Open Hypermedia. By using independent ontology, resource and metadata storage services, concepts referred to in Web resources can be identified and matched against potential ‘link destinations’ for navigational purposes. In the original DLS the ability to map from language terms to hypertext link destinations is governed by the existence of human-authored databases of hypertext links from which to choose. In COHSE, the process is driven instead by the various inter-relationships of concepts in the ontology.

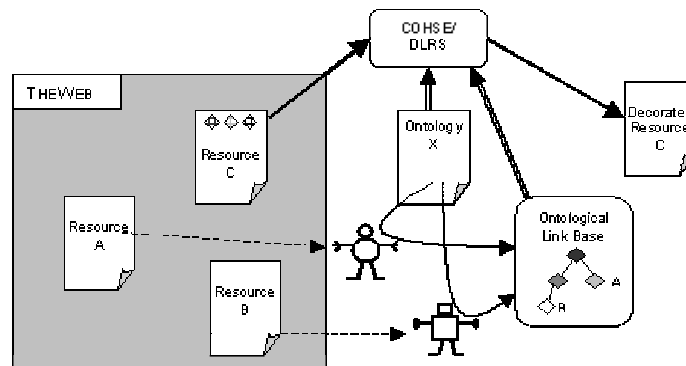


Figure 8: COHSE

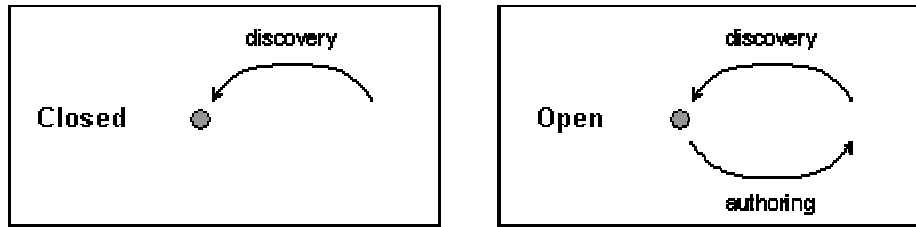


Figure 9: Adding Metadata

The COHSE prototype differs from systems such as RDF and SHOE in that its purpose is not to support query, but instead to provide extra information and linking for existing web pages. Note also that it's not confined to a particular set of web pages (e.g. those on a specific server) but any web page, so long as it has mechanisms for recognising the concepts in the documents. As it currently stands, this may involve simply matching words or terms given by the ontology, but is also extended to the addition of explicit metadata to the resources.

Discussion

Many of the systems described above focus primarily on metadata and its use in information discovery. COHSE intends to make use of metadata annotations in order to *build* and *construct* hypertexts. Here we discuss some of the issues that are of particular relevance to COHSE, in particular, the issue of discovery versus authoring. In the concluding discussion, we describe a space of metadata applications based on the richness of the model and degree of "openness" of the system.

By a *closed* system here, we mean one in which the links or associations between resources are "hard-wired" and fixed by the original author — for example using <A> tags in HTML. In contrast, an *open* system allows the addition of extra structure to resources (which may even be documents associated with third-party applications). This may not necessarily be through direct manipulation or amendment of the original source, but could be through the use of a proxy or similar that adds links or other metadata to the resources as they are read [20].

In general, approaches to metadata are about providing metadata for *resource discovery*. The resources are annotated with metadata describing their contents. The user (or an application) then queries this metadata in order to find resources, and is presented with some results.

In an Open Hypermedia framework, we are concerned not just with resource discovery, but also the authoring process — constructing a hypertext or links.

Adding metadata in a closed world manner (where some centralised repository of metadata exists which is queried in order to discover resources) is all about telling the user/application where to find resources, i.e. *resource discovery*. In the closed world, the metadata is an advert that tells other things what the resource is about and allows you to locate it.

Adding metadata in the OH framework is not only about resource discovery, but also describes how to link *from* the resource. In the open world, the metadata both advertises the resource and indicates where you can go from here.

Figure 9 shows the benefits of adding metadata in the two approaches. In the closed world, metadata provides links into the resource. In the open world, the metadata induces links both in and out of the resource. Of course, in the authoring model, resource discovery is still implicitly taking place, as the targets of the links must be obtained in some way from the associated

metadata. The COHSE prototype differs from other systems in that it's not so much to support query, but instead provides extra information and linking for existing web pages. Note also that it's not confined to a particular set of web pages. The link service will deal with any web page, so long as it has mechanisms for recognising the metadata in the documents. As it currently stands, this involves matching words or terms together with the recognition of explicit metadata to the resources. The ontology being used is a simple one (a thesaurus), but a richer model will be employed in the latter stages of the project.

Hypertext or Database?

Many approaches to conceptual metadata on the Web are concerned with resource discovery — how does the user (or an agent) locate the resources which are appropriate for their information need or task? The Web is being treated as a huge database rather than as a hypertext structure — links are, for the most part, ignored, and resources are located based on their content (or markup indicating their content). The line between linking and searching can become blurred [8], but in general the resource discovery approaches sit firmly at the searching end of the spectrum. In contrast to this, COHSE intends to address the problem of constructing hypertexts and building links, rather than simply providing resource discovery.

Metadata annotations can be used to either discover or author. During discovery, the system or agent locates resources according to their annotations. The results are often then simply presented as a ranked list — this is the "Web as database", with metadata used as an *indexing* mechanism. There is very little in the way of linking. If the metadata is used for authoring (as proposed in the COHSE approach), we are returning to the notion of the "Web as hypertext" — links are added to the source document as a result of the metadata annotations.

Q1) Given a resource or document, how do we determine the metadata ascribed to it?

Q2) Given some metadata, how do we determine the relevant resources?

In simple terms, we can consider **Q1** as being an indexing problem, while **Q2** is the process of retrieving resources given some index.

Other questions we may ask about an approach include:

- *Does it involve inserting metadata in the documents and if so, how is that achieved?*
- *Does it use a centralised repository of metadata, and if so, how is that constructed?*
- *What kind of representation is used?*
- *How are queries expressed?*

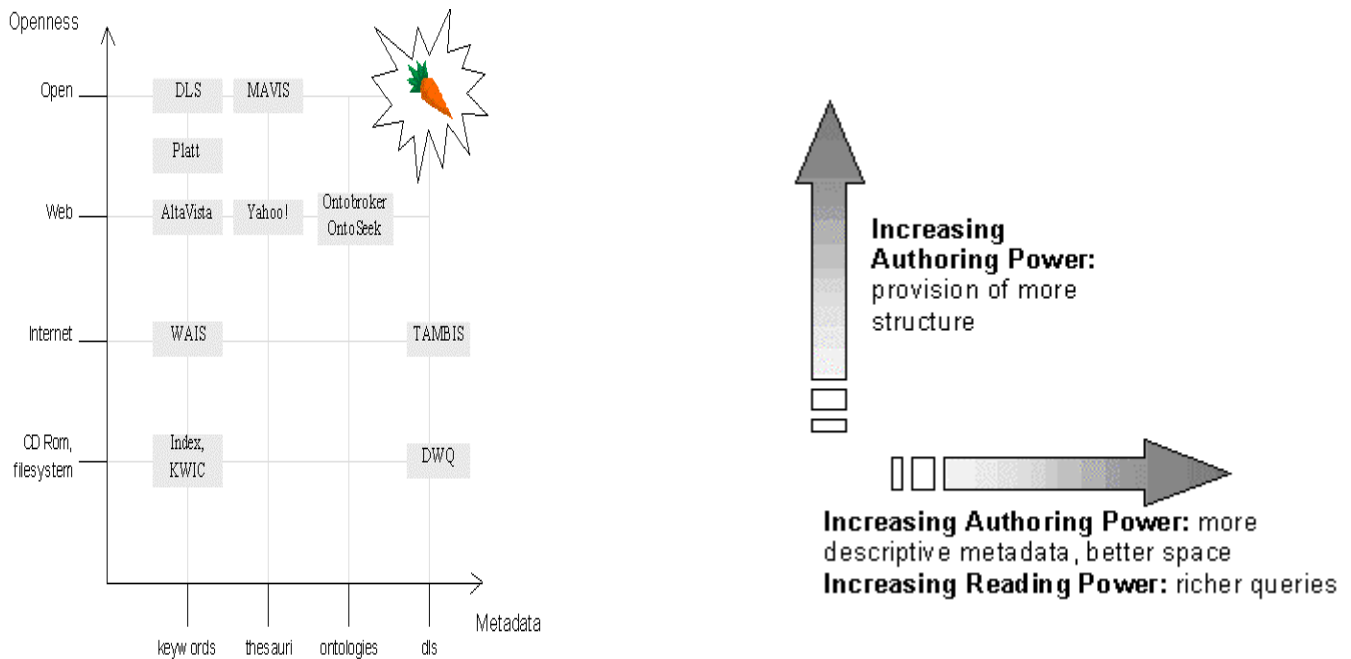


FIGURE 10: Metadata Space & Dimensions of Improvement

A Space of Metadata Approaches

In this section we focus on the issue of metadata annotation of web resources and the particular question of where the metadata resides and examine some existing approaches

We introduce a space that compares the “openness” of systems (in terms of distribution and the ease with which additional material can be incorporated) and a coarse notion of the expressiveness of the metadata. Figure 10 is an attempt to classify where different metadata approaches fit and how they relate to one another.

The x-axis represents increasing richness of the metadata representation, moving from keywords to thesauri, ontologies and finally some representation with an underlying reasoning representation (for example, a description logic). The y-axis represents an increasing use of openness or distribution or linking—the intention is that we move from simple standalone file systems to distribution via the net, the addition of linking on the Web, and finally open hypermedia.

For example, the Yahoo! classification uses a thesaurus (i.e. a static hierarchy of terms) in order to index web resources. The Platt demonstrator application [3] sits somewhere between the Web and the Open Hypermedia points as it uses links generated at runtime. However, the metadata model used is simple keywords. TAMBIS [1] is an example of a system that supplies access to distributed information sources on the net (giving its y-axis position), using a description logic model. The DWQ (Data Warehousing Quality) project [25] has also used a description logic to approach the problems of improving database schema integration. The Distributed Link Service (DLS) uses keywords as anchors when adding links to documents.

The dimensions of improvement that are gained through the addition of metadata and structure are also shown in figure 10. The increase in metadata richness allows the author to provide better descriptions of resources, and in turn enables the reader to pose richer queries which better meet the information need of the searcher. The addition of further structure increases the authoring power — in particular the jump to the Open Hypermedia architecture enables the use of link authoring through the provision of metadata. The Holy Grail (or in this case, carrot) for COHSE is the upper right hand corner of the graph — open hypermedia using a richly expressive model with reasoning.

Future Work

The ontologies that are used by the COHSE are a way of structuring a space of language terms. Consistent keyword descriptions are difficult to create and subsequently maintain, leading to an incoherent model of concepts and hence inaccurate linking, since each document will have many possible interpretations [6]. Consequently, some communities have developed domain-specific controlled vocabularies, or terminologies, based around a thesaurus of language terms, for example, The Art and Architecture Thesaurus (AAT) [30] or WordNet [18], a general language thesaurus based on semantic nets. These and others, for example, ACM Computing Classification System [11], are based on abstract hierarchical classification schemes and are used as classification schemes. However, their use as such is seriously hampered as they are largely predetermined, static and often unsound single classification hierarchies resembling “phrase books” [5]. They are not based on a systematic ontology - an explicit, rigorous, declarative specification of concepts - but are principled organisations of linguistic terms or phrases that do not have a

rigorous fixed interpretation other than that attributed to them through human interpretation. This lack of rigour makes them hard to: browse, use as a querying device, check for coherency, extend in a principled way and reason about [4].

Compositionally-based terminologies are much more powerful, resembling a collection of elementary concepts assembled according to formal rules to create composite concepts that are guaranteed meaningful. To be more effective such terminologies are best represented in a knowledge representation scheme that is expressive and can intrinsically support dynamic and automatic classification of complex composite concepts based on their components; such a scheme is a Description Logic (DL). Conventional frames or semantic networks do not have the logical concept subsumption and satisfiability reasoning services offered by DLs and are consequently less flexible when constructing and evolving the conceptual network and using it for retrieval. The next stage in the development of the ontology service is to incorporate a DL model, and to use that to control the update and maintenance of the ontology as it encounters new terms on the Web.

In various systems the conceptual model forming the links between documents is exposed and explicitly navigable [4, 28], whereas in [33] the classification scheme is more implicit. This raises issues of the presentation of links, the rendering of concept-based links, and the visibility of the ontology during linking and its use in query construction. Informed navigation implies that we inform the user of potential links between documents through shared or related concepts. Moreover, users can only search for terms in a controlled vocabulary if they know the vocabulary, and the controls and constraints upon it. Consequently, the visibility of the ontology during linking and its use in query construction is an issue under investigation.

Finally the editorial activity of the link service is to be strengthened so that the concept of a well-linked document is calculated across the whole document: linked concepts should be spread as evenly as possible throughout both the document space and the ontology space. Strategies for link culling and increasing are also being investigated: for example, is it valid to just choose on instance from each of the 'narrower terms' or is the effect of this kind of action too dependent on the design and construction of the ontology?

Concluding Remarks

The aim of a Conceptual Hypermedia Service is to escape from the limitations of a purely lexical string matching approach to link discovery. The prototype described here is successful in that the link construction software can interact with an independent suite of ontologically-motivated services. The fact that it can offer reasonable alternatives to the 'obvious' link candidates if the selection of links is unsuitable for some (editorially determined) reason demonstrates an advantage of the approach. Subsequent work will improve the user interface to the linking (ontological navigation) while improving the sophistication of the ontological processing and maintenance through the use of description logics.

Acknowledgements

This work was supported by EPSRC grant GR/M75426.

References

- 1 Baker, P. G., Brass, A., Bechhofer, S., Goble, C., Paton, N., and Stevens, R., TAMBIS: Transparent Access to Multiple Bioinformatics Information Sources. An Overview, in *Intelligent Systems for Molecular Biology ISMB98*. 1998: Montreal, Canada.
- 2 Barrett, R. and Maglio, P.P. (1998) Intermediaries: new places for producing and manipulating web content, In *Proceedings of Seventh WWW Conference*, Brisbane.
- 3 Bechhofer, S., Goble, C., and Drummond, N., Supporting Public Browsing of an Art Gallery Collections Database, in *11th International Conference and Workshop on Database and Expert Systems Applications DEXA 2000*, September 4-8. 2000: Greenwich, London, UK.
- 4 Bechhofer, S. and Goble, C. (1999). Classification Based Navigation for Picture Archives. To appear in *Proceedings of IFIP WG2.6 Conference on Data Semantics, DS8*, New Zealand, Kluwer.
- 5 Bechhofer, S., Goble, C.A., Rector, A.L., Solomon, W.D. (1997) Terminologies and Terminology Servers for Information Environments. Proc. IEEE Conf on Software Technology Experience & Practice, 484-497.
- 6 Bruza P.D. The modelling and retrieval of documents using index expressions. *SIGIR Forum* 25, 2 (1991), 91-103.
- 7 Bruza, P.D. (1990) Hyperindices: a novel aid for searching in hypermedia, In *Proceedings of the 1990 ACM Hypertext*, 109-122.
- 8 Carr, L., Links and Queries in the COHSE Project, 2000, Internal Project Report, available from <http://inanna.ecs.soton.ac.uk/cohse>
- 9 Carr, L., Davis, H., De Roure, D., Hall, W., Hill, G. (1996) Open Information Services, *Computer Networks and ISDN Systems*, 28 (7/11), 1027-1036, Elsevier.
- 10 Carr, L., De Roure, D., Hall, W., Hill, G., (1995) The Distributed Link Service: A Tool for Publishers, Authors and Readers, *World Wide Web Journal* 1(1), 647-656, O'Reilly & Associates.
- 11 Coulter, A. (1998) Computing Classification System 1998: Current Status and Future Maintenance. Report of the CCS Update Committee, *Computing Reviews*, Jan 1998, 1-5.
- 12 Crampes, M. and Ranwez, S., Ontology-Supported and Ontology-Driven Conceptual Navigation on the World Wide Web, in *11th ACM Hypertext Conference*, May 30 - June 4. 2000, ACM Press: San Antonio, Texas. p. 191-199.
- 13 Decker, S., Erdmann, M., Fensel, D., and Studer, R., Ontobroker in a Nutshell, in *Research and Advanced Technologies for Digital Libraries*. 1998, Springer Verlag LNCS 1513.
- 14 DeRose S., Maler, E., Orchard, D. and Trafford, B. *XML Linking Language (XLink) Version 1.0*, W3C Candidate Recommendation 3 July 2000. <http://www.w3c.org/TR/2000/CR-xlink-20000703/>
- 15 DeRoure, D., El-Beltagy, S., Gibbins, N., Carr, L. and Hall, W. (1999) Integrating Link Resolution Services using Query Routing, In *Proceedings of the Fifth Open Hypermedia Workshop* (in press).
- 16 DeRoure, D., L. Carr, W. Hall and G. Hill (1996) A Distributed Hypermedia Link Service, In *Proceedings SDNE96*, IEEE Computer Society Press.

- 17 Ellis, D., Furner, J., Willett, P. On the creation of hypertext links in full-text documents - measurement of retrieval effectiveness. *Journal of the American Society For Information Science*, 47(4) 1996, 287-300.
- 18 Fellbaum C. (ed.) (1998) *WordNet: An Electronic Lexical Database*, MIT Press, ISBN 0-262-06197-X
- 19 Fensel, D., Angele, J., Decker, S., Erdmann, M., Schnurr, H.-P., Staab, S., Studer, R., and Witt, A., On2broker: Semantic-based access to information sources at the WWW, in *World Conference on the WWW and Internet (WebNet99)*. 1999: Honolulu, Hawaii.
- 20 Gronbaek, K., Sloth, L. and Bouvin, N.O. (2000) Open Hypermedia as User-Controlled Metadata for the Web. In *Proceedings of 9th International World Wide Web Conference*. 553-566.
- 21 Grønbaek, K., Sloth, L. and Orbaek P. (1999) Webwise: Browser and Proxy Support for Open Hypermedia Structuring Mechanisms on the WWW. In *Proceedings of the Eighth International World Wide Web Conference*, 253-268.
- 22 Heflin, J., Hendler, J., and Luke, S., Coping with Changing Ontologies in a Distributed Environment, in *AAAI Conference Ontology Management Workshop*. 1999, AAAI Press. p. 74-79.
- 23 Heflin, J., Hendler, J., and Luke, S., SHOE: A Knowledge Representation Language for Internet Applications. 1999, Dept. of Computer Science, University of Maryland at College Park: Technical Report CS-TR-4078 (UMIACS TR-99-71).
- 24 Hitchcock, S., Carr, L., Jiao, Z., Bergmark, D., Hall, W., Lagoze C. and Harnad S., Developing services for open eprint archives: globalisation, integration and the impact of links. In *Proceedings of 5th ACM Conference on Digital Libraries*.
- 25 Jarke, M., Quix, C., Calvanese, D., Lenzerini, M., Franconi, E., Ligoudistiano, S., Vassiliadis, P., and Vassiliou, Y., Concept Based Design of Data Warehouses: The DWQ Demonstrators., in *ACM International Conference on Management of Data (SIGMOD'2000)*. 2000: Dallas TX, USA.
- 26 Lassila, O. and Swick, R. R., *Resource Description Framework (RDF) Model and Syntax Specification*. 1999, W3C: Proposed Recommendation (January 1999) <http://www.w3.org/TR/PR-rdf-syntax>
- 27 Lowe, D. and Hall, W. (1998) *Hypertext and the Web: An Engineering Approach* J. Wiley & Son.
- 28 Nanard J. & Nanard, M. (1991) Using structured types to incorporate knowledge in hypertext, In *Proceedings of the 1991 ACM Hypertext Conference*, 329-342.
- 29 Osterbye, K. and Wiil, U. (1996) The Flag Taxonomy of Open Hypermedia Systems, In *Proceedings of the 1996 ACM Hypertext Conference*, 129-139.
- 30 Peterson, T. (1994). *Introduction to the Art & Architecture Thesaurus*, OUP.
- 31 Thiele, H., The Dublin Core and Warwick Framework: A review of the literature March 1995 - September 1997. *D-Lib Magazine*, 1998.
- 32 Thistlethwaite, P. (1997) Automatic Construction and Management of Large Open Webs, *Information Processing and Management*, 33(2), 161-173.
- 33 Tudhope, D., Taylor, C. (1997) Navigation via similarity: automatic linking based on semantic closeness. *Information Processing & Management* 33, 2, 233-242.