Conceptual Open Hypermedia = The Semantic Web?

Carole Goble, Sean Bechhofer
Information Management Group
Department of Computer Science
University of Manchester
Oxford Road
Manchester M13 9PL, UK
+44 161 275 2000

{carole, seanb}@cs.man.ac.uk

Les Carr, David De Roure, Wendy Hall
Intelligence, Agents, Multimedia
Department of Electronics and Computer Science,
University of Southampton,
Southampton SO17 1BJ, UK
+44 23 8059 3255

{lac, dder, wh}@ecs.soton.ac.uk

ABSTRACT

The Semantic Web is still a *web*, a collection of linked nodes. Navigation of links is currently, and will remain for humans if not machines, a key mechanism for exploring the space. The Semantic Web is viewed by many as a knowledge base, a database or an indexed and searchable document collection; in the work discussed here we view it as a *hypertext*.

The aim of the COHSE project is to research into methods to improve significantly the quality, consistency and breadth of linking of Web documents at retrieval time (as readers browse the documents) and authoring time (as authors create the documents). The objective is *link creation* rather than resource discovery; in contrast, many existing projects are concerned primarily with the discovery of resources (reading), rather than the construction of hypertexts (authoring). The project plans to produce a COHSE (Conceptual Open Hypermedia ServicE) by integrating an ontological reasoning service with a Web-based Open Hypermedia link service. This will form a Conceptual Hypermedia system enabling documents to be linked via metadata describing their contents. The bringing together of Open Hypermedia and Ontology services can be seen as one particular implementation of the Semantic Web. Here we briefly present open and conceptual hypermedia, and introduce the architecture being employed within the COHSE project, and the prototype COHSE platform we have developed. We present the questions that we now plan to address that surround the Semantic Web when viewed from the perspective of a hypertext for people.

Keywords

open hypermedia, conceptual hypermedia, hypermedia authoring, ontology services, navigation

Semantic Web Workshop 2001 Hongkong, China

1. INTRODUCTION

The Semantic Web—a web of data defined and linked in such a way that its meaning is explicitly interpretable by software processes rather than just being implicitly interpretable by humans—is a vision held by many and articulated by Tim Berners Lee [5]. The issues surrounding the implementation of the Semantic Web are the focus of much current research. Much of this work, however, focuses on particular issues such as the implementation and representation of metadata, for example the OIL and DAML web languages [12, 15] or the activity of resource discovery, i.e. the finding of relevant information.

Less attention has been paid to the issues surrounding the delivery of this information once it has been located and the presentation of results, specifically to humans rather than software agents. Tim Berners-Lee's vision of the Semantic Web is that of a universe of network-accessible information, specifically:

- a means of people-to-people communication through shared knowledge, not just to browse but to create; and
- a space in which software agents can, through access to a vast amount of everything which is society, science and its problems, become tools to work with us. Machines can become capable of analysing all the data on the Web and collaborations will extend to computers.

In the recent enthusiasm to move the Web to one that has machine understandable semantics for automated processing we are in danger of forgetting that it has been successful not only as a machine readable infrastructure but also as one that is browsed by humans. In other words, the Semantic Web is still a *web*, a collection of linked nodes. Navigation of links is currently, and will remain for humans if not machines, a key mechanism for exploring the space. This should not be lost. The Semantic Web is viewed by many as a knowledge base, a database or an indexed and searchable document collection; in the work discussed here we view it as a *hypertext*.

The aim of COHSE is to research into methods to significantly improve the quality, consistency and breadth of linking of Web documents at retrieval time (as readers browse the documents) and authoring time (as authors create the documents). The objective is *link creation* rather than resource discovery; in contrast, many existing projects are concerned primarily with the discovery of resources (reading), rather than the construction of hypertexts (authoring). COHSE can be seen as a software agent that generates and presents links on behalf of both the author and the reader. Thus the Semantics of the Semantic Web are used for the automation of link generation, and the support of link navigation by humans.

The project plans to produce a COHSE (Conceptual Open Hypermedia ServicE) using three leading-edge technologies, two of which are drawn from the hypertext community:

1. an ontological reasoning service which is used to represent a sophisticated conceptual model of document terms and their relationships;
2. a Web-based Open Hypermedia link service that can offer a range of different link-providing facilities in a scalable and non-intrusive fashion;
3. the integration of the Ontology Service and the Open Hypermedia link service to form a Conceptual Hypermedia system to enable documents to be linked via metadata describing their contents.

COHSE is a system powered by OIL [15] (Ontology Inference Layer), of which Manchester is a leading developer. A prototype infrastructure has already been developed and demonstrated with a museum collection application provided by a local historical costume gallery.

The bringing together of Open Hypermedia and Ontology services can be seen as one particular implementation of the Semantic Web: the ontology service manages the Semantics and the open hypermedia manages the Web. It is, of course, not *the* implementation of the Semantic Web, as there are alternative architectures. In this position paper, we briefly present open hypermedia and conceptual hypermedia and introduce the architecture being employed within the COHSE project. We present the questions that we plan to address in phase two of the project now COHSE's current state provides a platform on which to experiment and explore the issues surrounding the Semantic Web from the perspective of a hypertext for people.

1.1 The COHSE technologies

COHSE builds upon three pillars: Ontology services for reasoning about metadata, Open Hypermedia and Conceptual Hypermedia. Here we briefly discuss these.

1.1.1 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 links from being abstracted from the documents and managed separately, as is made possible by XML's proposed XLink standard [18]. In *Open Hypermedia Systems* (OHS) links are first class objects, stored and managed separately from documents; 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 [20] and increasingly Web publishing applications adopt the open hypermedia approach [23].

The University of Southampton's Distributed Link Service (DLS) implements an open hypermedia system above the infrastructure of the World Wide Web [9, 10]. This provides a powerful framework to aid navigation and authoring and solves some of the issues of distributed information management [13]. Using an intermediary model [1], the DLS adds links and annotations into documents as they are delivered through a proxy from the original Web server to the ultimate client browser. It treats link creation and resolution as a service that may be provided by a number of link resolution engines that recognise different opportunities for adding various kinds of links to the documents. Thus the DLS creates a user-specific navigational overlay that can be used to superimpose a coherent interface to sets of unlinked or insular resources (such as the journal archives addressed by the Open Journals project [16]). Link resolutions include keyword recognition, the names of people and bibliographic citations as potential link anchors.

In the original DLS the link resolvers were hardwired into a monolithic system or chained sequentially. This inherently synchronous arrangement means that any delay was a delay in the critical path of document delivery, and hence all the processing was required to be relatively lightweight and tightly coupled.

In COHSE, the DLS was re-engineered to be a Distributed Link Resolution Service (DLRS) to allow link resolvers to be distributed across multiple servers and decoupled from the delivery of the document. Thus complex computation, such as ontological inferencing to add value for document authoring and browsing can be provided without impeding the delivery of the core document itself.

1.1.2 Conceptual Hypermedia Systems

To achieve the kind of diversity of association required for today's Web applications, documents need to be linked in many dimensions based on their content. Constructing such links manually is inconsistent and error-prone [14]. Furthermore, it obfuscates one of the chief reasons for associating documents; that their contents are *similar* in some way. *Conceptual Hypermedia Systems* (also known as Semantic Hypermedia Systems) specify the hypertext structure and behaviour in terms of a well-defined conceptual schema [7, 19, 24]. This types documents and links, and includes some kind of 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 [4]; concepts are linked and hence their associated documents are linked.

The University of Manchester's TourisT prototype experimented with a conceptual hypermedia approach for a Tourism Public Access System [8]. As the relationships between the concepts in the domain model evolve so do the links; as document concept descriptors change so do the links, making this a potentially powerful linking mechanism.

1.1.3 Ontology services for document metadata

Conceptual hypermedia presumes:

a) The description of nodes (web resources) that is precise and shared. For example all web resources that are about pets use the same terms and adhere to the same notion about what a pet is. This requires three things:

metadata: web resources are marked-up with descriptions of their content using a common syntax and model such as the Resource Description Framework, RDF [17]. However, marking up is no good unless everyone *speaks the same language*;

terminologies provide shared and common vocabularies of a domain, allowing search engines, agents, authors and users to communicate. RDF-Schema, (RDF(S)) [6], provides a standard way of defining standard vocabularies but doesn't actually define any. However, using a common vocabulary is no good unless everyone *means the same thing*;

ontologies provide a shared and common understanding of a domain that can be communicated across people and applications. Ontologies take a variety of forms, from hierarchical classification schemes such as the directories such as Yahoo!, thesauri, frame-based knowledge models, and logic-based models. Despite all these forms they all include a vocabulary of terms and some specification of the meaning of the terms.

b) A mechanism that can use those descriptions to infer new, previously undisclosed information about resources. Shared vocabularies based keyword collections do little to help here. Vocabularies based on ontologies that organize the terms in a form that has clear and explicit semantics can be reasoned over. For example, the metadata annotation of a web page states "this web page is about poodles”. An ontology states that poodles are kinds of dogs and dogs are kinds of pets. Thus we could infer that this page is also about pets and could be interesting to pet food retailers. Most ontologies have three major components that can be used in inference: a taxonomy, relationships between concepts and axioms (rules).

Representation languages for describing web resources and supporting inference over those resources, have been a major focus over the last year. Languages include RDF(S) [6], DAML [12] and OIL [15]. These last two have recently come together to form DAML+OIL.

OIL unifies the epistemologically rich modeling primitives of frames, the formal semantics and efficient reasoning support of description logics and a mapping to the standard Web metadata language proposals. Past work at Manchester has developed ontology servers that use description logics as a domain and metadata representation language [3]. The STARCH project has exploited the reasoning capabilities of description logics to describe and query the metadata of a collection of stock photography drawn from the Hulton-Getty collection of Getty Images [2]. As OIL is a variant on the description logic FaCT we have used previously, it was straightforward to adopt OIL as the COHSE representation language, and we already had the infrastructure to migrate our FaCT-driven ontology servers to OIL-driven ones.

2. COHSE ARCHITECTURE

COHSE provides a testbed for the exploration of the issues involved in building an open conceptual hypermedia system. The current prototype has four components, an ontology service, a resource service, a metadata service and a distributed link resolution service. Each of these components contributes an essential piece of the functionality required for an implementation of the Semantic Web. We do not propose COHSE as the definitive architecture for the Semantic Web, however, the components described here provide at least the minimum functionality that we consider is required.

In COHSE the **Distributed Link Resolution Service client agent** provides *presentation* and *delivery* of results, sometimes overlooked in favour of concentration on ontological matters and resource discovery. The DLRS uses a Concept Service to recognize concepts referred to in the web pages and uses these to link web pages together. The Concept Service is composed of two parts: the **Ontology Service** which provides *semantics* and world knowledge using a structured vocabulary; and the **Resource Service** which provides *resource discovery* mapping words or concepts to web pages. Figure 1 gives an overview of the architecture.

COHSE recognizes two sources of metadata describing a web page’s content:

1. Words or word combinations in the textual part of the resource itself, including XML tags, known to represent a concept. This is implicit ontology use as a word or phrase is used as a surrogate for a concept;
2. Concepts added explicitly through some metadata annotation process (automatic or manual) using OIL and referring to an OIL ontology. A resource may be described by multiple metadata descriptions and multiple ontologies. The metadata may be placed within the document itself (e.g. HTML `<META>` tags or an extended `<A>` tag) or in some external some metadata repository (e.g. an RDF repository or a linkbase). The **Metadata Service** manages this activity.

The metadata describing a document are the potential anchor points for links to web pages containing the same or *related* concepts. If concepts are identified through words, then these are the anchors. If they are explicitly marked up in the document, then the location of the markup, or the section of the document spanned by the markup is the anchor point. If the metadata is held outside the document then some indication of the scope of the anchor (a line, a paragraph, the whole document etc) is needed along with the concept.

2.1 Ontology Service

The Ontology Service maintains the ontology, and allows the application to interact with it through a well-defined API. Primarily, the Server provides operations relating to the content of the conceptual model. There are operations to extend the ontology, to query it by returning the parents or children of a concept, to return natural language terms for concepts and concepts for natural language terms, and to determine how concepts and roles can be combined to create new legal composite concepts. Concept requests are made in two forms, reflecting their occurrences in resources: words that present concepts and concepts expressed in OIL directly. A word necessitates an extra text-based resolution process to resolve the word to an appropriate concept. The Ontology Service manages multiple ontologies.

In the pilot the ontology is a simple hierarchical thesaurus providing a number of terms (concepts), along with a broader/narrower term hierarchy, related term links and synonyms. Although this is a simple model, the provision of hierarchical relationships has enabled us to conduct some initial experiments into the interaction between the DLRS and the Ontology Service.

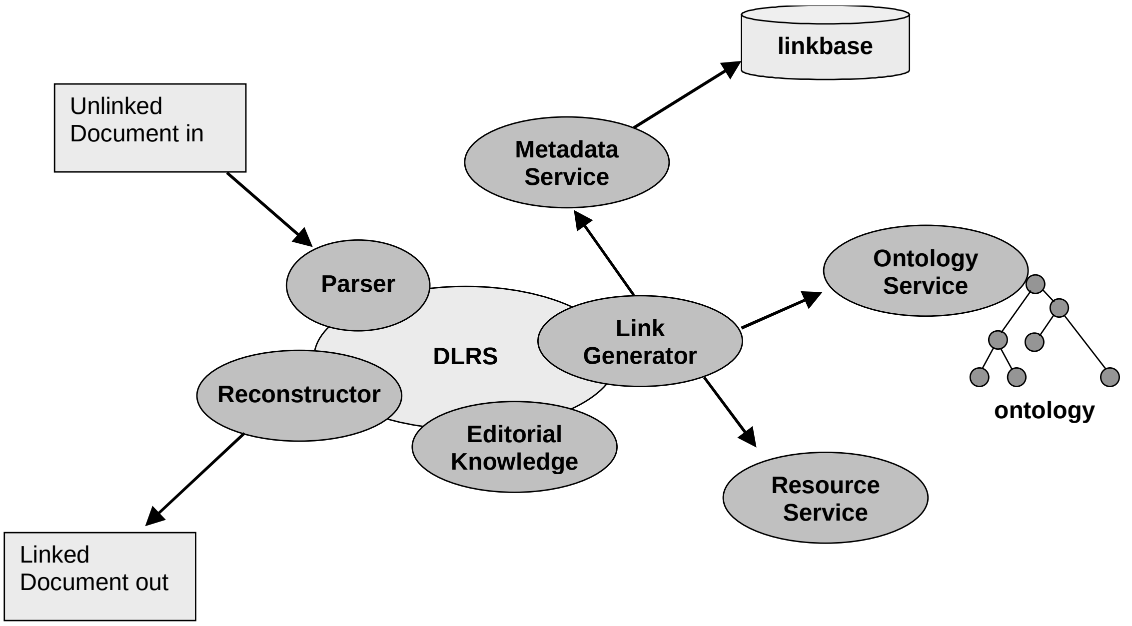


Figure 1 Architecture

The full Ontology Service uses ontologies represented in OIL. One of the benefits of OIL is that the expressivity of the model can be as simple or as complex as desired, and the migration of the simple thesauri to OIL is straightforward. More sophisticated expressive models benefit from the OIL's subsumption reasoning capabilities to use the ontology as a *semantic index* capable of supporting imprecise and abstract conceptual concepts, organising concepts in an inclusion classification scheme and supporting concept specialisation and generalisation.

2.2 Resource Service

The Resource Service maps concepts to resources.

2.3 Metadata Service

The Metadata Service allows documents to be decorated with metadata: concept expressions from a specific ontology. The service can either harvest specific tags from the documents themselves or apply external "metadata links" to read-only documents 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.

2.4 DLRS Client Agent

The DLRS Client Agent controls the user's interaction with the Ontology Service and Resource Service. It consists of two parts.

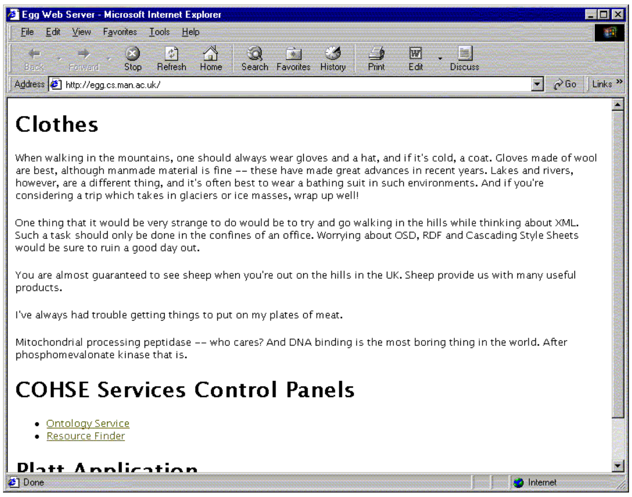
The low level component deals with interaction with the underlying representation of the document being viewed, i.e. the Document Object Model (DOM) object and interacts with the web browser. These are the "in and out" services of the DLRS.

The high level component is responsible for communication with the Ontology Service and Resource Service. This has two components: a link generator that promiscuously proposes linking opportunities and an editorial component that controls the linking opportunities by the application of filters, user profiles, similarity matrices etc, and the use of the ontology services provided by the ontology server. These are the "do it" services of the DLRS.

The DLRS agent receives the document and parses it into a DOM. The link generator examines the document, seeking linking opportunities within that document by examining the metadata associated with the document. For example, it contacts the Ontology Service for language terms that are used to represent the concepts in the ontology and recognizes a concept occurrence using the terms. It may also contact the Metadata Service for metadata held on the document in an external linkbase.

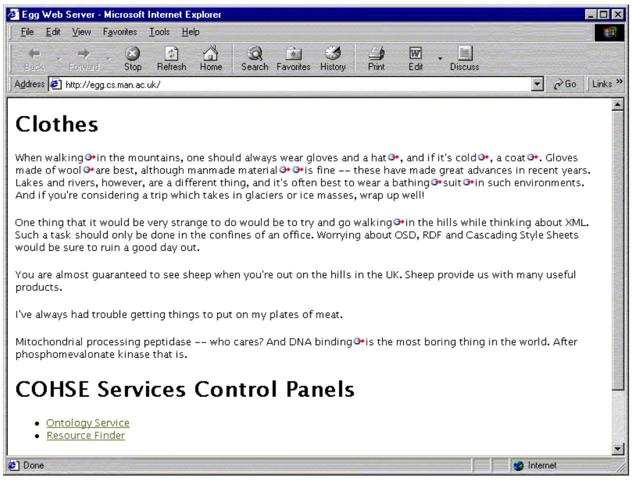
Figure 1 shows the architecture of the system. The *Ontology Service* manages the domain model of content and maps between natural language terms and the ontology. The *Resource Service* obtains web pages representing the concepts. The *Link generator* uses the ontology terms to make links. *Editorial knowledge* is used to prune or expand the links using the ontology

Having identified a concept, and hence determined a linking opportunity, the Resource Service is queried to retrieve possible destinations for the link, i.e. resources that contain instances of this concept. At this point, the DLRS may use ontological or semantic knowledge in order to make informed decisions about the links to choose. 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 web page*, it will choose to request more general or more specific concepts from the ontology service in order to expand or cull the set of anchor destinations.



Screen 1

Once all potential anchors have been explored, the reconstructor adds hypertext links with particular presentation styles and behaviours to the web page.

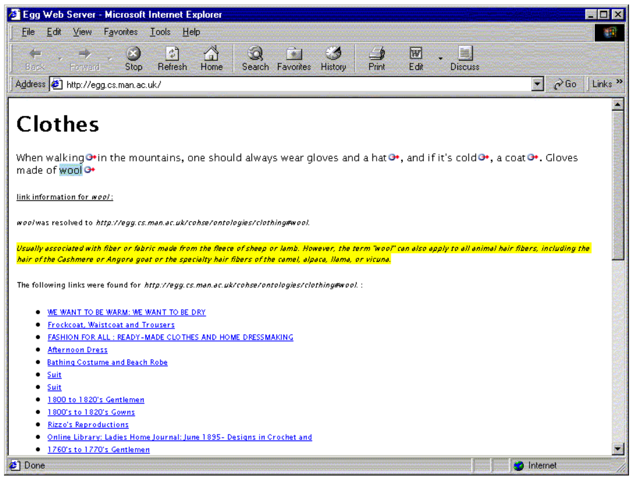


Screen 2

We developed a lightweight prototype COHSE for a costume museum application to see if the architecture was appropriate. In this pilot:

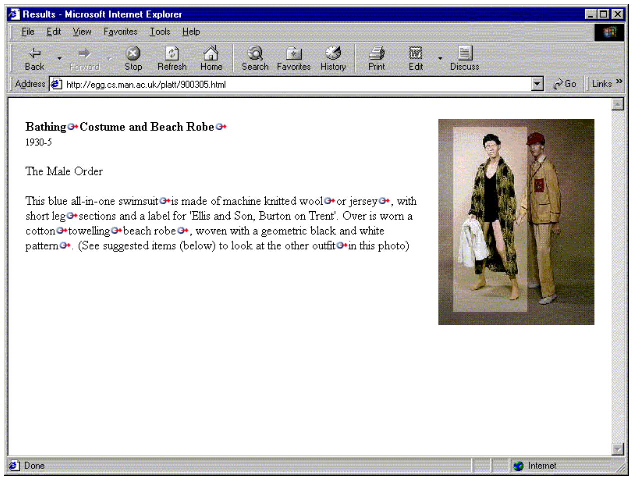
- the DLRS is integrated into the browser client, and consists of a Java applet that monitors the user's interaction with the browser together with a set of Javascript functions to manipulate the HTML DOM. The components of the link service are brought to bear on the web page as soon as it has been received by the browser. Once the set of links has been chosen the page is refreshed and redisplayed.
- the Ontology Service uses hierarchical thesauri of broader and narrower terms, related terms and synonyms. Queries and results are mediated through a simple XML document type.
- The metadata used to identify concepts in the documents are natural language terms in the web page.
- The Resource Service is a simple database mapping terms to web pages.

The screenshots show how links are added to an example document. The first shows a page about clothes. In the second, the page is shown linked against the clothing ontology (the small circle/arrow icons indicate anchor points).



Screen 3

The third screen illustrates the links available from the anchor concept "wool". Following the fifth link results in the browser refocusing to a page. Note that the ontology has been applied to this page, adding further links. In the next screen we can see that too many links were retrieved, so the editorial component has requested narrower (more specific terms) in a bid to reduce the number of links. The system is shown operating in debugging mode here, so all the initial links found are still being shown.

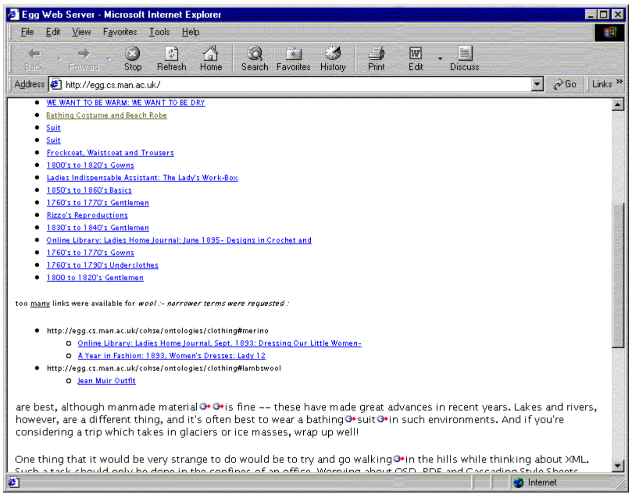


Screen 4

Finally, we see the effect of applying a different ontology (here a collection of geographical terms) to the initial page. By changing the link basis we have changed the hypertext.

2.5 Discussion

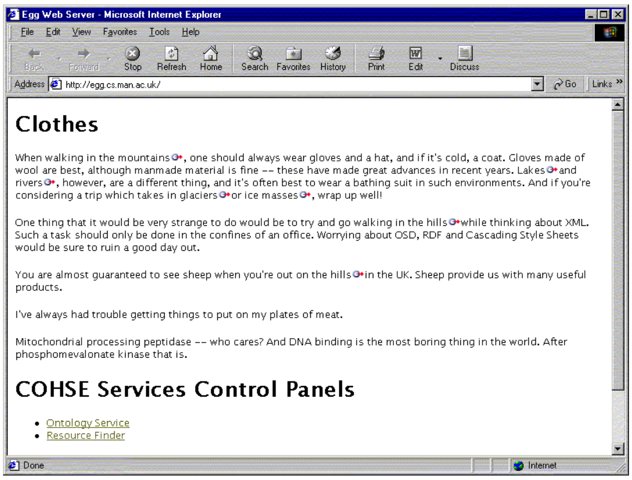
The intention of COHSE is to use metadata annotations to build and construct hypertexts. This is different from providing metadata for resource discovery, where an agent (person or machine) queries metadata in order to find some resources and is presented with a list of results. Here, the metadata's role is to advertise the resources content and allow others to locate it. Adding metadata in an Open Hypermedia framework not only describes how to link to a resource but how to link *from* a resource too. The metadata both advertises the resource and indicates where you can go from here, thus inducing links both in and out of the resource. The induction of links out of a resource is hypertext authoring. Of course,



Screen 5

resource discovery is still implicitly taking place as the targets for links must be obtained from the associated metadata.

From a hypertext authoring point of view, 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 more specific concepts the list of links can be usefully reduced whilst broadening the recognized concept can be used as a strategy to increase the number of links. However, how do we inform the user of potential links between documents through shared or related concepts? Currently links are listed,



Screen 6

but this looks like a search result. To look more like a hypertext perhaps only one candidate link should be presented, but then how to make that editorial choice? The ranking of concepts based on their similarity is tricky when concepts are defined by their descriptive properties as they can be in OIL [2]. Strategies for link culling and expansion need to be investigated that maintain the uniform appearance of the hypertext but do not preclude serendipitous link discovery.

Some Conceptual Hypermedia systems expose the ontology and make it explicitly navigable, e.g. [19]; others make the classification scheme more implicit [24]. Should the ontology be visible during linking? When the ontology becomes a sophisticated model of roles and axioms rather than a simple static tree, choosing a concept becomes an expression construction exercise.

The Karina project [11] uses descriptions (in a conceptual graph formalism) of resources to organize teaching materials in a hypertext setting. Resources are pulled together and presented in a ranked order (based on a number of factors such as prerequisites and time available). COHSE proposes a more general framework — source anchors and links can be being applied to *any* existing documents to build the hypertext.

Staab et. al. [21] discuss the construction of Community Web Portals using ontologies. The ontology is used to structure and present information and allows users to share a common language. A community web portal is, however, a very specific kind of structure in the Web, and this can be seen as more of a resource discovery than a link creation exercise.

Shah and Sheth [22] describe an approach using MREFs — Metadata REFerence links, which allow the conceptual markup of resources. What they decribe is still a *closed* system (in hypermedia terms) though, with the metadata marked up within the documents. Although the COHSE prototype discussed here uses words and markup within the document, the use of the open hypermedia pardigm within the general COHSE architecture will allow the separation of such metadata annotations from the resources, providing greater flexibility.

Other issues that we intend to address in the next phase of COHSE are: how can the reasoning services provided by OIL be best exploited by COHSE? does the combination of an Open Hypermedia framework and an OIL-based Ontology Services help to support change and evolution in the Semantic Web? The whole issue of how a person browses a hypertext that is generated "just in time" from metadata needs to be investigated. How should histories be maintained? What is the role of the back button? Bigger questions include: what is the role of document context and hypertext rhetoric in a COHSE, and is there a difference between querying and link following?

COHSE is now sufficiently developed to form a test-bed for a set of experiments to examine these questions, and to test the hypthosis that a Conceptual Open Hypertext = Semantic Web, or at least that Conceptual Open Hypertext $\subset$ Semantic Web.

2.6 Acknowledgements

The work was supported by EPSRC grant GR/M75426.

3. REFERENCES

[1] Barrett, R. and Maglio, P. P., *Intermediaries: new places for producing and manipulating web content*, in *Seventh WWW Conference*. 1998: Brisbane.

[2] Bechhofer, S. and Goble, C., *Classification Based Navigation for Picture Archives*, in *IFIP WG2.6 Conference on Data Semantics, DS8*. 1999, Kluwer Academic Publishing: Rotorua, New Zealand. p. 291--310.

[3] Bechhofer, S. K., Goble, C. A., Rector, A. L., Solomon, W. D., and Nowlan, W. A., *Terminologies and Terminology Servers for Information Environments*, in *Eighth International Workshop on Software Technology and Engineering Practice -- STEP97*. 1997, IEEE Computer Society: London. p. 484 -- 497.

[4] Beitner, N., Hall, W., and Goble, C. A., *Putting the media into hypermedia*, in *SPIE Multimedia Networking (1995)*. 1995.

[5] Berners-Lee, T., *Weaving the Web*. 1999: Orion Business.

[6] Brickley, D. and Guha, V. R., *Resource Description Framework Schema Specification 1.0*. 2000, WWW Consortium: W3C Candidate Recommendation www.w3.org/TR/rdf-schema.

[7] Bruza, P. D., *Hyperindices; a novel aid for searching in Hypermedia*, in *Proceedings of the ACM European Conference on Hypermedia Technology*. 1990. p. 109 -- 122.

[8] Bullock, J. and Goble, C., *TourisT: The Application of a Description Logic based Semantic Hypermedia System for Tourism*, in *Ninth ACM Hypertext Conference, Pittsburgh*. 1998.

[9] Carr, L., De Roure, D., Hall, W., and Hill, G., *The Distributed Link Service: A Tool for Publishers, Authors and Readers.* World Wide Web Journal, 1995. **1**(1): p. 647-656.

[10] Carr, L., Hall, W., and Hitchcock, S., *Link Services or Link Agents?*, in *Ninth ACM Hypertext Conference*. 1998, ACM Press: Pittsburgh. p. 113--122.

[11] 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.

[12] DAML, Darpa Agent Markup Language, http://www.daml.org.

[13] De Roure, D., Carr, L., Hall, W., and Hill, G., *A Distributed Hypermedia Link Service*, in *SDNE96*. 1996, IEEE Computer Society Press.

[14] Ellis, D., Furner, J., and Willett, P., *On the creation of hypertext links in full-text documents - measurement of retrieval effectiveness.* Journal of the American Society for Information Science (JASIS), 1996. **47**(4): p. 287--300.

[15] Fensel, D., Horrocks, I., van Harmelen, F., Decker, S., Erdmann, M., and Klein, M., *OIL in a Nutshell*, in *European Knowledge Acquisition Conference*. 2000, Springer-Verlag.

[16] Hitchcock, S., Quek, F., Carr, L., Hall, W., Witbrock, A., and Tarr, I., *Towards Universal Linking for Electronic Journals.* Serials review, 1998. **24**(1): p. 21--33.

[17] 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

[18] Maler, E. and De Rose, S., *XML Linking Language (XLink)*. 1998, WWW Consortium: Working Draft http://www.w3.org/TR/1998/WD-xlink-19980303.

[19] Nanard, J. and Nanard, M., *Using Structured Types to Incorporate Knowledge in Hypertext*, in *ACM Hypertext Conference*. 1991, ACM Press. p. 392--342.

[20] Osterbye, K. and Wiil, U., *The Flag Taxonomy of Open Hypermedia Systems*, in *ACM Hypertext Conference*. 1996. p. 129--139.

[21] S.Staab. J.Angele, S.Decker, M.Erdmann. A.Hotho, A.Maedche H.-P.Schnurr R.Studer ,Y.Sure *Semantic Community Web Portals* in Proceedings WWW9, the ninth Conference on the World Wide Web, Amsterdam 2000

[22] K. Shah and A. Sheth *Logical Information Modelling of Web-accessible Heterogeneous Digital Assets* in Proceedings of ADL 98, Advances in Digital Libraries, Santa Barbara, CA, 1998

[23] Thistlethwaite, P., *Automatic Construction and Management of Large Open Webs.* Information Processing &Management, 1997. **33**(2): p. 161--173.

[24] Tudhope, D. and Taylor, C., *Navigation via similarity: automatic linking based on semantic closeness.* Information Processing & Management, 1997. **33**(2): p. 233--242.