Modelling and Visualizing Perspectives in Internet Digital Libraries

Simon Buckingham Shum, Enrico Motta & John Domingue
Knowledge Media Institute, The Open University, Milton Keynes, MK7 6AA, U.K.

Email: sbs@acm.org, E.Motta@open.ac.uk, J.B.Domingue@open.ac.uk
WWW: http://kmi.open.ac.uk

Abstract
This paper is concerned with the interpretation of scholarly documents. We begin by examining core tasks that face a scholar in interpreting a scholarly document, or a collection of documents. We then argue that current approaches to document description, and current technological infrastructures, particularly over the World Wide Web, provide very poor support for these interpretive tasks. We describe an approach to modelling the perspectives in which documents are embedded, based on researchers’ claims about their own work, and those of others. We detail how such a model is being implemented in a Web environment, and the services it could provide to a scholarly community. These include the creation, visualization and interpretation of conceptual structures reflecting the relations between different research efforts, scholarly perspectives and debates.

1. Introduction
The mere publication of information does not constitute a body of knowledge; nor does simply obtaining information constitute understanding. Obtaining documents is just the first step; meaning and significance arise through their interpretation, which results in an understanding of the perspective adopted, and possibly other perspectives. In this paper we describe an approach to modelling and visualizing perspectives in internet digital libraries. We begin by examining core tasks that face scholars in interpreting documents, and collections of documents (§2). We argue that current approaches to document description, and current technological infrastructures, particularly over the World Wide Web, provide very poor support for these interpretive tasks (§3). We describe an approach to modelling the perspective in which documents are embedded, based on researchers’ claims about their own work, and those of others (§4). We detail how such a model is being implemented in a Web environment (§5), and the services it could provide to a scholarly community. These include the creation, visualization and interpretation of conceptual structures reflecting the relations between different research efforts, scholarly perspectives and debates, which we exemplify with a worked example (§6). We then discuss related work, key issues that this work raises, and the next steps in our research programme (§7).

2. The scholarly work of interpretation
Contextualising ideas in relation to the literature is a fundamental task for authors and readers—are they new, significant, and trustworthy? Scholars accomplish this firstly by bringing to bear their own knowledge of the field. This then leads to commentary and discourse of various kinds, which reflect the extent to which peers regard an author’s work as authoritative, from private annotation of a document, to formal peer review of conference/journal submissions, to published reviews of literatures and books. In the context of annotation and peer review, we have described elsewhere a publishing toolkit (D3E) that converts a scholarly document into a structured discussion website, and an electronic journal (JIME) which uses this to support an innovative peer review model. These are all processes that already take place in the world of paper, although we have argued elsewhere that the electronic medium has particular properties that can transform peer review in important ways (Buckingham Shum and Sumner, 1998; Sumner and Buckingham Shum, 1998). We can think of conventional scholarly publication and debate as a document-centred, text-based process. Text is a rich medium in which to publish and discuss ideas in detail and with subtle nuances, but the corresponding disadvantage is that it takes a long time to read, and is hard to analyse computationally.

A complementary approach described in this paper focuses on the conceptual models implicit in textual documents and discourse. The goal is to provide a summary representation of ideas and their
interconnections, in order to assist literature-wide analysis. We propose that this has advantages over textual media for tracing the intellectual lineage of a document’s ideas, and for assessing the subsequent impact of those ideas, that is, how they have been challenged, supported and appropriated by others. In addition, the availability of explicit conceptual models opens possibilities for automatic analysis of a community’s collective knowledge.

We begin with the idea that an author’s goal is to persuade the reader to accept his/her perspective, which constitutes a set of claims about the world. Usually, the author has some new ideas that s/he is contributing, and asserts particular relationships between these and existing ideas already published in order to demonstrate both the reliability of the conceptual foundation on which s/he is building, and the innovation and significance of the new ideas. The scholarly reader’s task is to understand which ideas are being claimed as new, and assess their significance and reliability.

Let us switch from a reading scenario to the scenario of literature search and analysis. In this case, the scholar has some ideas and relationships in mind that s/he is trying to locate in the literature—has anyone written about them, or perhaps these ideas exist but not yet in a single document? The interpretive task includes formulating the ideas of interest in a variety of ways that may uncover relevant documents, reading the documents (as just described), and then interpreting the documents to characterise any patterns that appear to emerge. This is a similar scenario to that of a newcomer to a scholarly community (e.g. a student; librarian; lecturer or researcher from another discipline) who wants to know, for instance, what the seminal papers are, or if there are distinctive perspectives on problems or classes of technique that define that community.

We contend that scholars are very poorly supported in these tasks by conventional library and technological environments, but that digital libraries open up new possibilities which have yet to be exploited. Our approach is to analyse the tasks that scholars bring to documents and libraries, and then design appropriate support.

Consider the document interpretation scenario. In the non-digital world, there is currently no way beyond following citations (only those provided by the author), or using citation indices (to find others citing him for some reason), to ask questions such as:

- “Has anyone built on the ideas in this paper, and in what way?”
- “Has anyone challenged this paper?”
- “Has anyone proposed a similar solution but from a different theoretical perspective?”

Considering the literature analysis scenario, there is currently no way for a scholar to query a digital library with analytical/interpretive questions of the following sorts:

- “Are there any documents building on theory T, but which contradict each other’s predictions?”
- “Are there any documents applying method M to domains D and E?”
- “Is there any Web-based software which tackles problem P?”
- “What impact did Theory T have?”
- “Are there distinctive theoretical perspectives on problem P?”

These are arguably the kinds of phenomena of most interest to scholars when they write papers, engage in debate or search the literature. These are also the kinds of questions asked by researchers unfamiliar with a literature, including students. We now describe an approach that holds the potential to provide support in identifying significant conceptual structures such as these.

3. Evolving the Web beyond simple linking

The World Wide Web is the first global hypertext system to emerge, providing a rudimentary infrastructure for publishing interlinked documents and discourses. This level of representational sophistication is already extremely useful for enhancing, even transforming, scholarly publication and discourse, as we have
However, the Web provides little support for structuring, searching or analysing scholarly concepts, documents or discourses. Early, pre-Web hypertext systems have already demonstrated (on a small scale) the power of features such as semantically typed nodes and links, bidirectional links, composite nodes which represent more complex structures, and structural searching. It is increasingly recognised that the Web would benefit from such features (see for instance the analysis of ‘fourth generation’ Web functionality by Bieber et al. (Bieber et al., 1997)).

Information retrieval using statistical and text analysis techniques include techniques for clustering and mapping documents based on semantic similarity (Chen and Carr, 1999; Chen and Czerwinski, 1998), and for inferring certain types of inter-document relationships (e.g. “cites”, or “summarises”) (Allan, 1996). Automatic techniques clearly have the advantage that they can be applied to large text corpora with little human effort once the texts are in a suitable format for processing. From the perspective of scholarly interpretation, the key weakness of such techniques is that it is extremely hard, if not impossible, to automatically identify more complex kinds of scholarly relationships between documents such as those given above. Human-encoded document descriptions are required to express such scholarly claims and structures.

We propose that when a new article is ready for dissemination, authors describe the document’s main contributions and relationships to the literature using a controlled vocabulary analogous to a metadata scheme (but implemented using a formal ontology), and submit the description to a networked repository. We describe in the next section the concept of an ontology, and the representational scheme underpinning our approach.

4. Representing scholarly claims

4.1 Ontologies

An ontology in philosophy refers to a model of what exists in the world (Bunge, 1977). The artificial intelligence community has appropriated the term to mean the construction of knowledge models (Gruber, 1993; Newell, 1982) which specify concepts or objects, their attributes, and inter-relationships. A knowledge model is a specification of a domain, or problem solving behavior, which abstracts from implementation-centered considerations and focuses instead on the concepts, relations and reasoning steps characterizing the phenomenon under investigation. Our application of knowledge modelling in this project is to implement a semantic network which expresses important aspects of the web of ideas and perspectives implicit in the documents and minds of a scholarly community.

4.2 An ontology for representing scholarly claims

An ontology reflects a (typically community-wide) viewpoint on how best to conceptualise a particular domain or phenomenon. Hence, its main role is to support knowledge sharing and reuse. It might appear paradoxical, therefore, to propose the use of ontologies to support scholarly communities in managing their knowledge, since conflicting worldviews, evidence and frames of reference lie at the heart of research and debate.

The key issue is in what is represented. It is hard to envisage when scholarly communities will no longer need to make claims about, or contest, the nature of a document’s contributions (e.g. “this is a new theory, model, notation, software, evidence”), or its relationships to other documents (e.g. “it applies, modifies, predicts, refutes...”). Our approach builds on this relatively stable dimension of what are otherwise constantly evolving research fields, by representing scholars’ claims about the significance of ideas and concepts—a focus on discourse and argumentation (how scholars support and contest claims), and on context (the conceptual network in which an idea is embedded). Representing concepts separately from claims about them is critical to supporting multiple perspectives.

We are adopting a philosophy of ‘minimal ontological commitment’ (Gruber, 1995) and incremental formalisation (Shipman and McCall, 1994), which reflects an emphasis on making explicit just enough structure to be usefully expressive and enable the provision of valuable computational services, but leaving the document texts to express the details and nuances of an author’s argument (as opposed to trying to
formalise it). This minimises the effort required to submit a document description; if there is evidence that authors wish to link ontological concepts to specific paragraphs within a document (e.g. as proposed by Heflin et al.), then we can provide a way to do so, but we will begin at the document level. A scheme suitable for a wide range of disciplines is proposed in Figure 1, but it is both generalisable and tailorable to other fields (e.g. a more experimental field might specialise Idea into Hypothesis; another might not need Software).

![Figure 1: Representational elements for making claims about the key contributions of a publication, and relationships to other concepts.](image)

We hypothesise that a given community should be able to agree on a relatively small set of uncontroversial conceptual and relational types which can adequately express the majority of claims made. The goal is to design an ontology which is simple enough to understand without being simplistic, yet expressive enough that most researchers can represent the key claims made in most documents.

This approach tackles the persistent problem facing any ontology development effort, namely, that the world being described is typically dynamic, necessitating resource intensive updating and restructuring. Shifting the representational focus to the claims that researchers make about contributions (node types) and their relationships to the literature (link types), avoids the problem of committing to conceptualisations that may be invalid in 2-3 year’s time.

### 4.3 Representing scholarly argumentation and perspectives

A fundamental requirement for our system is that it must be able to represent multiple scholarly perspectives. ‘Perspectives’ exist at many levels, starting from arguments about specific issues (e.g. did X cause Y?; was this a good experimental design?), which cumulatively develop into contrasting theoretical perspectives and paradigms with associated terminologies, methodologies, and criteria for judging validity. Our own and others’ research into argumentation support tools demonstrates the importance of not overloading untrained users with a complex argumentation scheme that takes a lot of effort to use ‘correctly’ (Buckingham Shum, 1996; Buckingham Shum et al., 1997; Marshall and Shipman, 1995; Selvin, 1999; Shipman and Marshall, 1999). Thus, in keeping with our philosophy of incremental formalisation and minimal ontological commitment, the proposed ontology can be used to make contributions to an argumentation structure initially at the level of supports, raises issues with, and refutes, to make it as easy as possible for a scholar to add an argumentation link to a concept or document in the literature. Many more elaborate argumentation schemes have been proposed (Lee, 1990; Schuler and Smith, 1990; Smolensky et al., 1988; Toulmin, 1958), but our analysis of this literature shows little evidence of successful uptake (Buckingham Shum and Hammond, 1994). Our conclusions are that more elaborate schemes should be introduced only when there is the demand from a user community. Rigorous and carefully maintained argumentation networks make many kinds of useful analysis possible, but make assumptions about users’ expertise and consistency of representation that are unlikely to hold in the context of an open, internet community. The details of an author’s reasoning are therefore left to the document’s text, and are not made explicit in the knowledge base.
5. Implementation

5.1 Knowledge modelling infrastructure

Our approach relies on a suite of knowledge modelling technologies developed at the Knowledge Media Institute over a number of years. The OCML modelling language (Motta, 1998) supports the construction of knowledge models by means of several types of construct. It allows the specification and operationalization of functions, relations, classes, instances and rules. It also includes mechanisms for defining ontologies and problem solving methods (Benjamins and Fensel, 1998b), the main technologies developed in the knowledge modelling area. Problem solving methods are specifications of reusable problem solving behaviours. OCML has been used in several projects, in domains such as medicine, geology, engineering design and organizational learning. As a result the language is now associated with a large library of reusable models, providing a useful resource for the knowledge modelling community. In our scenario, OCML provides the formalism for defining our ontology for scholarly debate and interpretation, henceforth referred to as ScholOnto.

WebOnto (Domingue, 1998) is a web based tool which enables knowledge engineers to collaboratively browse and edit knowledge models over the Web. The architecture is composed of a central server and a Java applet. WebOnto’s central server is built on top of a customised web server LispWeb (Riva and Ramoni, 1996) and uses OCML as the underlying modelling language. In addition to implementing the standard HTTP protocol, the LispWeb server offers a library of high-level Lisp functions to dynamically generate HTML pages, a facility for dynamically creating image maps, and a server-to-server communication method. The WebOnto Java applet provides multiple visualizations of OCML knowledge models, a direct manipulation and forms interface for creating new knowledge structures, and a groupware facility which supports both synchronous and asynchronous model building by teams of knowledge engineers (illustrated in Figure 2 and Figure 9).

Applied to the problem of managing scholarly knowledge concepts and documents in online research communities, these technologies provide the building blocks for a scaleable Web infrastructure.

5.2 Ontology design

Figure 2 shows the top level structure of the ontology, as specified in OCML. Both nodes and links in the semantic network created by scholars’ submissions are SCHOLARLY-KNOWLEDGE-CONCEPTS. Nodes are SCHOLARLY-CONTRIBUTION-ELEMENTS, and links SCHOLARLY-RELATIONSHIPS, which are subdivided into ARGUMENTATION-LINKs and NON-ARGUMENTATION-LINKs.
Figure 2: Main taxonomy of the ScholOnto ontology, also illustrating the WebOnto ontology browser and editor.

Figure 3 shows the class definitions for the scholarly concepts of SOFTWARE, METHODOLOGY and LANGUAGE. SOFTWARE is defined as a SCHOLARLY-CONTRIBUTION-ELEMENT, which ADDRESSES PROBLEMS, USES/APPLIES any other kind of SCHOLARLY-CONTRIBUTION-ELEMENT (e.g. a METHOD or LANGUAGE), and MODIFIES/EXTENDS other kinds of SOFTWARE. METHODOLOGY and LANGUAGE are similarly defined.

```lisp
(def-class SOFTWARE (scholarly-contribution-element)
  "This class models all the software tools which are used in the course of research. We try to model here the relations between a piece of software and the various concepts, models, theories, etc which are implied by a particular approach."
  ((addresses :type problem)
   (uses-applies :type scholarly-contribution-element)
   (modifies-extends :type software)))

(def-class METHODOLOGY (scholarly-contribution-element)
  ((addresses :type problem)
   (modifies-extends :type methodology)
   (uses-applies :type scholarly-contribution-element)))

(def-class LANGUAGE (scholarly-contribution-element)
  ((addresses :type problem)
   (uses-applies :type language :type theory-model)
   (modifies-extends :type language)))
```

Figure 3: OCML definitions of SOFTWARE, METHODOLOGY and LANGUAGE (see main text)

The ontology is designed to support scholars in making claims by asserting relationships between concepts. Other scholars may support, raise-issues-with, or refute these claims. Figure 4 shows schematically the structure of a scholarly “claim” in the ontology.
The OCML specification associated with the structure in Figure 4 is shown in Figure 5. A claim is formally defined as a relation between a set of authors, who make a legal-scholarly-assertion, with some justification. A legal-scholarly-assertion is a statement instantiating a scholarly-relationship (e.g., addresses, predicts, refutes) between two elements (e.g., methodology X addresses problem Y). The justification is free text supporting a claim. An author will not be expected to enter this since the justification for their claim is already to be found in the document they are describing. But if another scholar, for example, supports, raises-issues-with, or refutes another’s claim (see scheme in Figure 1), without publishing an associated document, then some form of textual justification is expected. This could in turn point to a more rigorous justification in another document (ideally, directly accessible).

To both support and refute a particular claim is inconsistent. If an author is a member of two sets of authors who have made such conflicting assertions (which can be automatically checked), then we can say that his/her position with respect to a particular assertion is inconsistent (Figure 6).

The design of the ontology is based on the analysis of scholarly articles from a range of different fields, taking about two person weeks’ effort. Once the top level structure stabilised, it required only two days to develop the first version of the ontology in OCML. Preliminary testing of the ontology indicates that it is already adequate to support the scenarios as described in this paper.

6. Revisiting the scenarios

We now return to our scenarios of literature and document interpretation, and use a worked example to clarify how our tools could support scholarly work. Within the hypertext research literature, one of the landmark articles is the summary of the Dexter Hypertext Reference Model by Halasz and Schwartz (Halasz and Schwartz, 1994), which specifies both semiformally, and formally (using the Z notation), abstract
properties of hypertext systems, enabling comparison of existing systems, and specification of theoretically possible future systems. Figure 7 shows a prototype user interface for submitting the article’s description to the repository.

Figure 7: Prototype forms-based interface for submitting the description of claims in a document.
The user interface guides users through the schema using dynamic menus, and enables them to browse and search for existing concepts to assist their reuse. Some domain concepts are simple to reference (e.g. the name of a specific software system, framework or methodology), whilst others are less concrete and could benefit from information retrieval support, e.g. finding the name(s) used to describe a domain problem (“user disorientation”), an idea (“a global hypertext system”), or an empirical phenomenon such as a piece of evidence (“low ability students benefit most from physics simulations”).

This would generate an OCML entry in the ontology, as shown in Figure 8.

```
(def-instance dexter-htxt-ref-model-article article
  ((describes-scholarly-contribution-element dexter-ht-ref-model
    (concerns-domain hypertext-hypermedia)
    (has-author halasz-f schwartz-m)
    (has-title "The Dexter Hypertext Reference Model")
    (publication-details "Communications of the ACM, 37 (2), 30-39")
    (has-url "www.acm.org/pubs/articles/journals/cacm/1994-37-2/p30-halasz/"
      (acm-ccs "I.7.2" "H.1.1" "H2.1" "H3.2" "H5.1")
    ))
)

(def-instance dexter-ht-ref-model theory-model
  (addresses absence-of-standards
    (analyses notecards augment concordia hypercard hyperties intermedia kms-zog
      neptune-ham)
    (predicts theoretically-possible-system-in-dexter-ht-ref-model)
    (uses-applies Z)))
```

Figure 8: The OCML entry for the Dexter article, declaring its contributions to the literature (dexter-ht-ref-model, which is a theory-model, and predicts theoretically possible systems), and its relationship to other concepts (analyses several existing systems, and uses-applies the Z notation).

The article is now added to the ScholOnto knowledge base, enabling users to ask questions such as, “What motivated the Dexter Hypertext Reference Model, and what impact has it had?” A forms-based interface, generated automatically from the ScholOnto ontology by WebOnto, enables users to ask such questions through simple menu selection (Figure 9). In the longer term, we can envisage adding a natural language interface to our set of tools.
Figure 9: A forms-based interface generated automatically from the ScholOnto ontology by WebOnto enables users to query the model via menu selection. The screenshot shows several possible queries to analyse the motivation behind, and impact of, the Dexter Hypertext Model (we have combined them to save space; in reality one would most likely submit these as separate queries). The queries specify, respectively, (1) interest in the theory-model: dexter-hypertxt-ref-model, (2) what problems does it analyse?, (3) are there any theory-models which modify-extend it?, and (4) is there any software which uses-applies it?

The knowledge base could generate visualizations (Figure 10) showing the Dexter Model’s motivation and conceptual roots (links to the left), and the nature of the work which has built on it since by the respective authors, or other researchers (links to the right).
Figure 10: Representational schema which will provide the basis for generating visualizations of the literature. An author has described (links to the left) the relationships that motivate and situate an article’s key contribution (a Reference Model, marked by the central node) within its literature. Subsequent researchers (links to the right) have modified/extended the model, and implemented software systems based on their extended models (e.g. the “DeVise” system, lower right).

7. Discussion

In this section we describe some of the technical and research possibilities opened up by the design we have presented, and contextualise our approach to related research.

7.1 Intelligent services

A knowledge model enables inference-based searching and alerting. It will be possible to ask the system questions such as “What impact did Theory T have?”, since “impact” can be defined, for example, in terms of the number of subsequent documents using or modifying it, the number of different domains in which it has been applied, the number of problems addressed which drew on the theory, and so forth. Our knowledge modelling environment makes it simple for us (as system maintainers) to write heuristics that could assist in finding relevant documents, e.g. “if Method Y extends Method X, and Method X is challenged, then Method Y may be challenged”. We will also assist scholars in composing their own rule-based interest profiles, e.g. “If 3 or more documents support Language L and challenge Language M (or any Languages based on them), and 3 or more documents support Language M and challenge Language L, then send me a concept map showing their interconnections” —since this may be evidence of two schools of thought.

Another important advantage of our approach is that the existence of a formally represented knowledge model makes it possible to envisage additional reasoning services on top of the ‘basic’ search support. For instance, it will be possible to develop specialized agents whose goal is to identify emerging perspectives, using heuristic knowledge and machine learning techniques. For instance, an agent could discover a ‘European perspective’ on a particular issue, by analyzing the geographic spread of the relevant positions. As these machine-discovered assertions are added to the knowledge model, software agents effectively become actors in the scholarly debate. This scenario raises interesting issues, both from a social and a research point
of view, and it is an example of the general trend towards reducing the boundaries between humans and machines (Stutt and Motta, 1998).

7.2 Minimising the cognitive effort to encode metadata
We aim to test the hypothesis that lowering the effort required to add new metadata to a library significantly improves its level of uptake. We have a large bibliographic database of hypermedia publications which we hope to offer immediately as a searchable digital library. Knowing that one’s publications are already described in the library may encourage scholars to annotate their entries with the new metadata required to generate the concept models and maps. We will contrast this repository with a research community for which we do not have a bibliographic database to start with.

7.3 Scholars are not librarians
Internet-based digital libraries of the sort that concern this project will change the roles of librarian and scholarly researcher established by paper-based, geographically-based libraries. For an internet-based library to scale realistically, with potentially tens of submissions arriving every day, the only people who can be expected to initially describe new articles are the people with most motivation to maximise the visibility and impact of the work—the authors. However, authors are not librarians or knowledge engineers who traditionally have possessed the skills to do information classification. This raises two issues: whether scholars are able to describe their documents sufficiently well to enable the system to make use of their descriptions, and how to make the underlying description technologies accessible and understandable. We follow initiatives to develop metadata schemes for the Web (see next section) in assuming that given intuitive representation schemes and user interfaces, domain experts will be able to submit useful descriptions of their own work. These are, however, empirical questions that will be addressed through lab-based studies studying detailed interaction with the system, followed by broader field trials once the system is deployed in different research communities.

7.4 Relationships to other research
Our technologies can be seen as a conceptual and technical development from current efforts to develop metadata description schemes for the Web. Metadata in the context of digital libraries refers to ways to encode information about resources in machine-interpretable formats, typically by completing a standard set of descriptive fields. Well known examples include USMARC (LoC, 1994) for library resources, Dublin Core (Dublin Core, 1995) to provide a simple high level scheme for web resources, and IMS (IMS, 1994) for educational resources. The W3C also has a metadata working group (W3C, 1997a) overseeing Web-specific initiatives, and the proposed Resource Description Framework (W3C, 1997b) for interrelating multiple metadata schemes refers to the need for knowledge modelling concepts similar to those presented here. However, in all of these initiatives, the goals are to describe the content of resource sites/documents, with comparatively little emphasis given to relational metadata. This contrasts with our goal of representing multiple claims about content, where there is no single right answer.

From a representational and technical perspective, our approach differs from metadata in that ontologies support more sophisticated modelling, for example, specifying sufficient and necessary conditions for relations (e.g. Figure 5), and providing metalevel modelling support which makes it possible to reason about the ontology itself. OCML in particular also provides powerful inference support making it possible to directly operationalise the ontology (ScholOnto), and its instantiation (the ScholOnto knowledge base).

The (KA)² initiative (Knowledge Annotation for Knowledge Acquisition) (Benjamins and Fensel, 1998a) aims to support the knowledge acquisition community in building a knowledge base of its own research by populating a shared ontology. The knowledge base is constructed by authors annotating their web pages (e.g. publications; personal and project pages) with tags (analogous to HTML META tags), which can be read by a specialised search engine called Ontobroker (Fensel et al., 1998). The key architectural difference to our approach is that (KA)² semantic tags are embedded in the physical content, whereas our approach decouples content from claims about its status. This architectural difference reflects the different aims of the two enterprises. The aim of (KA)² is to capture the contents of web pages in a formalism which can be reasoned about by Ontobroker. In our scenario we do not aim to represent directly the domain-specific content of a paper, but the debate about the scholarly status of that content. Moreover, we are concerned that authors will
not be prepared to invest the required effort to encode models of document content, and as argued earlier, we cannot assume a stable ontology to describe an active research field.

Finally, ontologies are beginning to be used in the context of digital libraries, although for different purposes to those set out here. Ontologies can assist the extraction of concepts from unstructured textual documents (Embley et al., 1998), by serving as a source of knowledge about the particular topic. Ontologies can also assist in managing document descriptions in large digital libraries (Weinstein, 1998; Weinstein and Alloway, 1997).

8. Conclusion

The Web has established itself within many research communities as a medium for rapid document publication and access. However, its support for many of the other tasks that scholars need to perform is extremely weak, despite the fact that semantic hypertext systems, as the Web was originally envisaged by Berners-Lee, are well suited to tasks such as structural searching, pattern analysis, and heuristic filtering.

Our approach is novel in that the proposed ontology expresses scholarly claims about domain concepts, not the domain concepts themselves. Our aim is to support scholarly debate, and the creation of author-centred and community-wide perspectives by scholars or software agents. We focus on the conceptual models implicit in textual documents and discourse in order to provide a summary representation of ideas and their interconnections. This has advantages over textual media for tracing the intellectual lineage of a document’s ideas, and for assessing the subsequent impact of those ideas. In addition, the availability of explicit conceptual models opens possibilities for automatic analysis of a community’s collective knowledge. There is no comparable work within the metadata or knowledge modelling communities, whose primary focus is on describing the content of resources, rather than discourse about content.

9. References


D3E: Digital Document Discourse Environment (D3E). Knowledge Media Institute, Open University. <d3e.open.ac.uk>


Dublin Core: Metadata Element Set (1995)


JIME: Journal of Interactive Media in Education. Knowledge Media Institute, Open University. <www-jime.open.ac.uk>


