Available online at www.sciencedirect.com
SCIENCE DIRECT
Int. J. Human-Computer Studies 64 (2006) 420–445
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Sensemaking tools for understanding research literatures: Design, implementation and user evaluation

Victoria Uren*, Simon Buckingham Shum, Michelle Bachler, Gangmin Li[1]

Knowledge Media Institute, The Open University, Milton Keynes MK7 6AA, UK

Received 25 October 2004; received in revised form 25 July 2005; accepted 18 September 2005
Available online 16 November 2005
Communicated by E. Motta

Abstract

This paper describes the work undertaken in the Scholarly Ontologies Project. The aim of the project has been to develop a computational approach to support scholarly sensemaking, through interpretation and argumentation, enabling researchers to make claims: to describe and debate their view of a document's key contributions and relationships to the literature. The project has investigated the technicalities and practicalities of capturing conceptual relations, within and between conventional documents in terms of abstract ontological structures. In this way, we have developed a new kind of index to distributed digital library systems. This paper reports a case study undertaken to test the sensemaking tools developed by the Scholarly Ontologies project. The tools used were ClaiMapper, which allows the user to sketch argument maps of individual papers and their connections, ClaiMaker, a server on which such models can be stored and saved, which provides interpretative services to assist the querying of argument maps across multiple papers and ClaimFinder, a novice interface to the search services in ClaiMaker.

Keywords: Modelling interfaces; Search interfaces; User studies

1. Introduction

Researchers are benefiting from improved access to documents through digital libraries, electronic journals, eprint archives, etc., but improved access brings its own problems. There is less time to track the growing numbers of conferences, journals and reports they can access. Researchers are interested in questions such as: How does the expert community perceive this theory, model, language, empirical result? Where did this idea come from? What kind of evidence supports it and challenges it? Are there different schools of thought on this issue? Answers to these kinds of questions arise out of the private sensemaking activity which is integral to reading the literature. By 'sensemaking' we refer to Weick's (1996) work on how individuals and groups construct meaning when confronted by complex, sometimes contradictory information. We literally 'make sense' by giving form to our evolving understanding of the meaning of data and ideas, as we seek to relate them to our existing conceptual structures, through writing, talking, sketching and other forms of external representation. In the absence of a single canonical view of the world, we must construct 'plausible narratives' to fill in the gaps. Within scholarly discourse, there are accepted ways of establishing (and contesting) plausibility. In this kind of sensemaking, past reading assists the interpretation of related documents which in turn lead the reader on to explore new texts.

In this paper, we describe the prototype tools which have been developed to support sensemaking and report on a case study in which the tools were put to use. In Section 2, we present the aims of the Scholarly Ontologies Project, for which the tools were developed and outline its approach to representing scholarly argument. In Section 3, we present related work. In Section 4, we describe the tools,

*Corresponding author. Tel.: +44 1908 858516; fax: +44 1908 653169.
E-mail addresses: v.s.uren@open.ac.uk (V. Uren), s.buckingham.shum@open.ac.uk (S. Buckingham Shum), m.s.bachler@open.ac.uk (M. Bachler), ggl@omii.ac.uk (G. Li).
[1]Present address: Open Middleware Infrastructure Institute, School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK.

1071-5819/$ - see front matter
doi:10.1016/j.ijhcs.2005.09.004

ClaiMapper, ClaiMaker and ClaimFinder. In Section 5, we report on the case study. This study had two parts. In the first, ClaiMapper was used to construct a model of a literature as a network of claims. This part was about the use of the tools to record and support sensemaking. In the second part of the study, ClaiMaker and ClaimFinder were used to examine the claim network. This part looked at whether the models could be usefully interpreted by users other than the creator. In Section 6 we reflect on the advantages and limitations of the approach.

2. The Scholarly Ontologies Project

The Scholarly Ontologies Project was an EPSRC project, funded as part of the programme on Distributed Information Management, with the aim of developing a 'claims server' to support scholarly interpretation and argumentation. It investigated the practicality of publishing explicit, semi-formal conceptual structures in a collective knowledge base. These structures are grounded in conventional documents which are accessible, via hyperlinks, directly from the claims server. In this way, the claims server has the additional role of a new kind of index to distributed digital library collections. The system enables researchers to make claims: to describe and debate their view of a document's key contributions and relationships to the literature.

2.1. Representing interpretation and argument

"Ontology" is the term used in knowledge modelling to describe an abstract specification of concepts, attributes and relationships whose meanings are agreed by the ontology's users (Gruber, 1993). Typically, ontologies are applied to control interpretation or semantic annotation in a specific domain, such as travel, enabling interoperability between, for instance, airline and hotel booking sites.

In contrast, we propose an ontology not only to represent consensus, but also principled *dis*agreement, which can support multiple interpretations. These might be different interpretations of the claims in a single paper or between a number of papers. The resolution of the apparent contradiction between the use of ontologies to control interpretation and our use of an ontology to represent multiple interpretations comes from the observation that, while researchers do not necessarily agree on the issues under debate, the mechanisms of scholarly debate do remain stable over time. Whether research is in the arts or sciences, there will always be problems that are of key interest, people will put forward theories, predictions, hypotheses, etc., and try to support them with data and analysis. These contributions may, in their turn, be challenged or developed further. In order to tackle the problem of multiple interpretations, our knowledge modelling effort has focused on capturing these enduring, discipline-independent relationships between objects, which we call discourse relations, rather than the types of objects. This and other requirements for the ontology required for representing scholarly debate in a claims server are outlined in Table 1.

The base form of the representation is a directed graph in which Concepts form the nodes and the links are drawn from a taxonomy of discourse relations. Concepts are stored as short pieces of free text succinctly summarising a contribution, at whatever granularity the researcher wishes. A claim is a triple of two such objects connected by a link (Fig. 1). Other objects which may be used as nodes in claims include sets (collections of concepts) and claims themselves.

Within the ontology we have a taxonomy of link types to represent the rhetoric of researchers when they present their arguments (see Fig. 2). Relations are classified into groups with similar rhetorical implications: Supports/Challenges, Problem Related, Taxonomic, Causal, Similarity and General. Each relation belongs to exactly one of these groups. Some of these groups, such as Supports/Challenges and Problem Related enable the user to take positions. Others, particularly the Causal and Taxonomic relations, support the building of models of domains for arguments to refer to. These are not argumentation relations. However, we discovered that they were necessary to allow users to provide supporting material to make their arguments comprehensible. It can be argued that these two categories of groups should be split at a higher level of the taxonomy into relations about content and relations about positions. However we have not followed this route in the version presented here. The design and evolution of the ontology is described by Buckingham Shum et al. (2002). Its theoretical relationship to discourse relations theory, specifically Cognitive Coherence Relations theory, is detailed in Mancini (2005).[2]

Each relation is identified by a natural language label. This can be changed for communities with different rhetorical styles helping us tackle requirements 1 and 6 in Table 1.

Each relation is assigned two properties: a polarity which indicates whether it has positive or negative implications (e.g. the label *proves* has positive polarity whereas *refutes* has negative polarity; it implies *dis*proof) and a weight (high or low) which indicates how forceful it is (e.g. *refutes* is more forceful than *disagreesWith*). The assignment of polarity allows us to tackle requirement 2 in Table 1. The assignment of polarity and weighting is illustrated for the Supports/Challenges class in Table 2.

In addition to Concepts, two other kinds of object can be used as the nodes in Claims. These are Sets, groups of concepts brought together by the user because they share a

[2]Cognitive Coherence Relations theory is derived from research into coherence relations in text and speech. Approaches in these fields such as Discourse Representation Theory (DRT) (Kamp, 1981) focus on modelling formally sentential relationships and sub-structure, but this is too fine a granularity to expect from users except trained discourse analysts. Approaches such as DRT might be used to analyse the content of concepts in claims.

Table 1
Requirements for the ontology

Requirements for the discourse ontology
1. *Mimic natural language* expressions to *reduce the cognitive gap*: The underlying structure should be based on a noun/verb metaphor with the relations taking the role of verbs. Making arguments in pseudo-natural language was intended to make the scheme intuitive for contributors.
2. *Permit the* expression *of dissent*: A classical truth maintenance model would not be fit for purpose; if ''truth'' is established on an issue, it ceases to be worth doing research about. The scheme must therefore be closer to that presented by Toulmin (1958), with evidence being presented in favour of claims and complemented by counter-claims. To support or challenge a claim, the modeller uses relations with either positive or negative polarity. The concept of polarity is drawn from the work of Knott and Mellish on Cognitive Coherence Relations (CCR) (Knott and Mellish, 1996). To agree with a proposition, the relation used would have positive polarity. To disagree it would have negative polarity. Giving relations polarity opens up the possibility of providing services at a higher level of granularity than that of individual link labels. See Mancini and Buckingham Shum (2004) and Uren et al. (2004) for further discussion on CCR.
3. *Signal the ownership of public content*: Contributors must take responsibility for the claims they make because we depend on the social control of peer pressure to motivate high quality claim making. Although a peer-reviewing process could be conceived we have not attempted one in the early prototypes. Ownership also has a key role in the claims server as digital library server: claims would be ''backed up'' by a link to a published paper. There is an analogy here with Toulmin's warrants.
4. *Accommodate the social dimensions to being explicit*: Argument modelling invites researchers to consider making explicit what is normally implicit in the text of a paper (discussed in Buckingham Shum et al., 2000). Consider a relation *refutes*. This is a forceful term and therefore should carry greater weight in calculations than, for example, *takes issue with*. From the social side, some contributors might prefer to use the less extreme term when linking to concepts created by eminent figures. Providing these soft options recognises the social dimensions to citation, and aims to remove a possible barrier to adoption.
5. *Assign concepts no category outside of use*: We require that the typing of object should be optional and that objects may change their type depending on the context. A key precept of conventional approaches to ontologies is that objects in a scheme are typed under one or more classes. While this is acceptable for non-controversial attributes such as *Software*, this cannot be sustained when we are talking about the role that a concept plays in multiple arguments. The concept that is a *Problem* under debate in one paper may be an *Assumption* in another (or even within the same paper). The scheme must therefore allow the same concept to take on several types in different situations.
6. *Assist in making connections across disciplinary boundaries*: We are trying to identify a core set of argumentation relations that are useful in many disciplines. However, the precise terms used for making a case will differ from one research community to another. We tackle this using the idea of dialects. Drawing again on Cognitive Coherence Relations (see 2.), we define a core set of relational classes, with properties such as type, polarity and weight, but these may be reified with natural language labels in many ways. For instance, a community in which it would be strange or unacceptable to *refute* someone's work could change the label to something they felt more comfortable with (e.g. *raises serious questions*), but the basic properties of the strongly negative relation that challenges a concept would remain unchanged. This method would let us configure claim servers for different communities without altering the underlying reasoning engine.

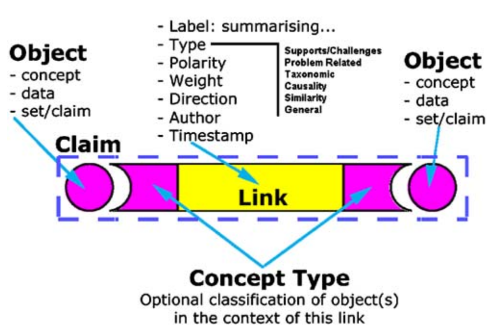


Fig. 1. Structure of a claim.

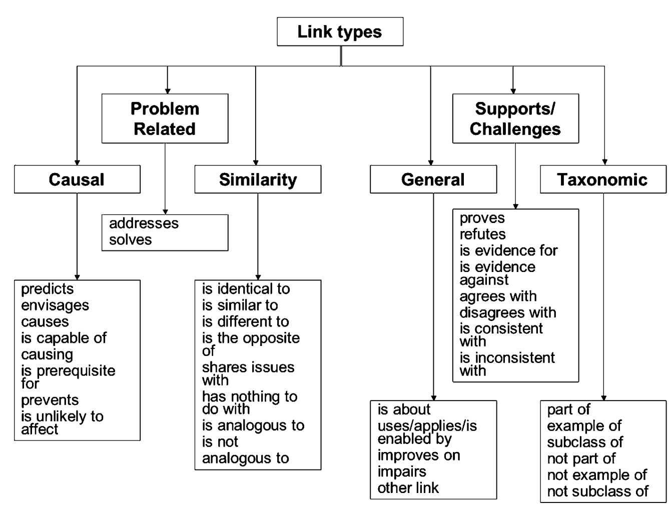


Fig. 2. Taxonomy of rhetorical link types.

common theme and Claim triples themselves. This single level of nesting allowed users to build complex arguments, while mitigating the implementation difficulties posed by fully recursive structures. Later in the prototyping cycle an extension of Sets was implemented which allows an element of a Set to be another Set or a claim. This allows more deeply nested structures to be built.

A concept may be assigned a type (e.g. ⟨Data⟩ ⟨Evidence⟩ ⟨Hypothesis⟩). However typing of concepts is optional (although in a pedagogical context it might be appropriate to enforce concept typing as a way to lead students to think more carefully about their claims). Optional typing is unusual in ontologies but is motivated

Table 2
Summary of discourse ontology parameters for the Supports/Challenges class of links

Label	Polarity	Weight
Proves	Positive	High
Refutes	Negative	High
Is evidence for	Positive	Low
Is evidence against	Negative	Low
Agrees with	Positive	Low
Disagrees with	Negative	Low
Is consistent with	Positive	Low
Is inconsistent with	Negative	Low

by the observation, made in requirement 5 of Table 1, that objects may play different *roles* in different contexts, since researchers may disagree on the node's type: e.g. is this *Language* also a *Theory*? Is this based on *Opinion* or *Data*? One person's underlying *Theory* may be someone else's *Problem*. An author may even problematize an idea later on in a text, which she has up to that point treated as an unproblematic *Theory*, *Language*, etc. Our approach, therefore, is that if a type is assigned it is stored as part of the link connecting it to another concept rather than as an intrinsic part of the concept. This has the effect of making node-typing context dependent and thus permits multiple typing of the same concept in different relations. We also argue that for some relations typing the concepts is redundant since the role of the nodes is implied by the relation type. The link types *is evidence for*, which implies the left-hand node is a piece of evidence and *addresses*, which implies the right-hand side is a problem and the left a remedy, are examples of this behaviour.

The cost of this approach is that removing any compulsion on users to type their concepts does reduce the level of detail of analysis possible. Here we chose to trade off computational power against cognitive load on users and produced a "good enough" representation. We envisage that certain communities might want to complement the discourse ontology with their own specialist vocabularies, particularly where these are widely used; for example, a biomedical version might need to type concepts that are genes, enzymes, proteins, etc. However domain specific extensions of this type were not attempted in the generic prototype discussed in this paper.

To illustrate discourse ontology links in use, we will examine some claim triples based on arguments presented in Borodin et al.'s paper "Finding Authorities and Hubs From Link Structures on the World Wide Web".[3] These claims were produced by the first author during the modeling phase of the case study reported in Section 5 and are also presented in Fig. 3.

[3]Borodin, A., Roberts, G.O., Rosenthal, J.S., Tsaparas, P., 2001. Finding authorities and hubs from link structures on the World Wide Web. In: Proceedings of the 10th International Conference World Wide Web Conference (WWW10), Hong Kong.

2.1.1. *"TKC effect—algorithm favours tight knit communities" "is about" "link ranking algorithms"*

The first claim triple (A in Fig. 3) uses a link with type General and label *is about*. It expresses a topic membership relation. This sort of relationship tends to be indicated quite early in papers when authors are indicating the domain they are addressing. For example, the abstract of Borodin et al.'s paper starts with the sentence "*Recently, there have been a number of algorithms proposed for analyzing hypertext link structure so as to determine the best "authorities" for a given topic or query*". They are helpful in Claim networks as the topic node gives an entry point for browsing.

2.1.2. *"TKC effect—algorithm favours tight knit communities" "is different to" "SALSA behaviour—algorithm mixes authorities from different communities"*

The second triple (B in Fig. 3) uses a link labeled *is different to* with type Similarity and negative polarity. It expresses a negative similarity relation, i.e. it says that "TKC effect" and "SALSA behaviour" are not the same. In the original paper, one of the places where this claim is expressed reads as follows: "*Specifically, when computing the top authorities, the Kleinberg algorithm tends to concentrate on a "tightly knit community" of nodes (the TKC effect), while SALSA tends to mix the authorities of different communities in the top authorities*".

2.1.3. *"TKC effect—algorithm favours tight knit communities" "is capable of causing" "topic drift"*

The third triple (C in Fig. 3) uses a link with type Causal and the label *is capable of causing*. One of the sources for this claim in the paper reads: "*…these examples seem indicative of the topic drift potential of the principal eigenvector in the Kleinberg algorithm*". This is an example of how the discourse ontology constrains the modeler, here to make a claim which is rather stronger than that in the original paper. How should we read causal statements in Scholonto? We have stated that we wish to deal with models of the arguments people make, rather than propositions about the world (see requirement 2). Yet this claim looks on the surface like a proposition that could take a truth-value. However, if we add in the metadata stored for claims about the creator and backing paper, the claim could be read as: Uren states, that Kleinberg claims, that "TKC effect—algorithm favours tight knit communities" *is capable of causing* "topic drift". The claim as a whole is able to be used as a node within other claims; for example, a claim about a (hypothetical) rebuttal made by Kleinberg or another reader/modeller's interpretation of the same text.

Fig. 3 shows the three claims above in the context of the claim network in which they were created. The arrangement is dominated by the claim highlighted in the centre of the model: "TKC effect—algorithm favours tightly knit communities" *is different to* "SALSA behaviour—algorithm mixes authorities from different communities". This

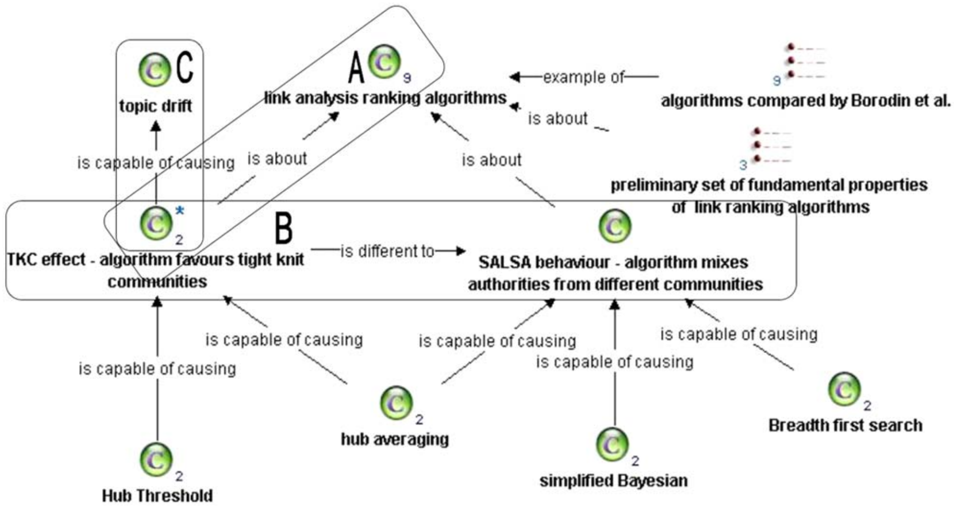


Fig. 3. ClaiMapper model of the paper by Borodin et al. The figure uses the icon conventions from ClaiMapper, which will be described in full in Section 4. These are all basic links with the orb icon representing a concept and the arrow a (labeled) relation.

claim summarizes the two key phenomena explored in Borodin's paper. Below this line are arranged the algorithms investigated, each linked to the phenomenon or phenomena it can cause. Above this line the more general information is placed, these links set the work in context and link one of the phenomena, causally, to a problem, labeled "topic drift". Within the private world of ClaiMapper, the relative position of concepts can be used in this way to assist sensemaking interaction with the paper. However it is not transferred to the server, which stores only connectivity information since it merges models from many papers and many users which may have overlaps.

3. Related work

3.1. Hypertext

It is widely recognized that hypertext was shaped by the Memex vision of Bush (1945) and the NLS system of Engelbart (1962). It is less commonly known that both of these pioneers saw the construction and analysis of scholarly arguments as key applications of their technologies, as discussed in Buckingham Shum (2003). The work described here can thus be traced back to Bush and Engelbart, via the extensive research in the 1980s and early 1990s into hypertext graphical argumentation tools and more recent work on scholarly hypermedia (for a review see Buckingham Shum, 2005). Bush proposed the idea of "associative trails", or chains of documents linked by associations similar to the associations in human memory. We propose ScholOnto claim networks as a method of signposting these trails through a document collection such as the Internet. This progression from the closed pre-Web argumentation systems to the Internet increases the scale of the user community proportionally.

3.2. Semantic annotation

Recent years have seen the early stages of the development of the Semantic Web, in which web pages with machine interpretable mark up provide the source material with which agents and semantic web services operate (Berners-Lee et al., 2001). The commentary offered by the ScholOnto approach could be viewed as a form of semantic annotation of documents. The W3C annotation project Annotea (Kahan et al., 2001) and CREAM (Handschuh and Staab, 2002), an annotation framework being developed at the University of Karlsruhe, offer alternative infrastructures for managing mark-up of this kind. Annotea applies the W3C open standards for annotating XML and HTML documents and assumes the Web as the environment. CREAM applies the same standards but is aimed a knowledge management environment, such as company intranets, where a lot of data may be stored in databases or other non-web-native forms and where more control of annotation quality may be desired (and possible).

3.3. Concept mapping

The use of familiar metaphors is essential when presenting new technologies to users to reduce the barrier to uptake. In designing the representational scheme for the Scholarly Ontologies project we sought familiar sensemaking methods that link ideas. The Mind Maps developed by Buzan (1989) are well established as a sensemaking method in education and business. Mind Maps typically have the main topic in the centre of the map with subtopics and sub-subtopics radiating off like seeds on the head of a dandelion. However the focus on a central concept is too restricting a practice for sensemaking in research literatures, where people may explore several inter-related topics and so the fundamentally hierarchical Mind Map method would not be appropriate. Therefore we took an approach which is more akin to the concept mapping method proposed by Novak and Gowin (1984). In concept mapping the concepts are linked into networks, rather than hierarchies, giving more freedom to express the interrelations between ideas. The notion of labelling links, which is widely used in concept mapping, is also important. However, because we wanted to be able to construct services using our links, we needed to have a degree of control over the kinds of links modellers would use and support the reuse of structures via a server.

3.4. Link vocabularies

In developing this vocabulary of link types we were aware of a trade-off between usability and expressiveness. Users may be wary of a very complex system, preferring not to use it rather than be seen to make a "mistake". On the other hand very simple systems impose too many constraints on the types of models users can build. Therefore while we drew on theoretical work such as that on Cognitive Coherence Relations (Knott and Mellish, 1996; Knott and Sanders, 1998) we took a pragmatic approach to selecting relations, aiming for a moderate palette of useful links rather than a very complete set such as that proposed by Trigg (1983). Since then Gil and Ratnakar have published work on the TRELLIS system (Gil and Ratnakar, 2002) which also uses discourse relations. We note that they too selected relations that could be understood by users, rather than "precise" or "complete" relations. Selected examples of discourse relations used in Trellis include: provides background for, in contrast with, is solved by, is motivation for, depends on and causes.

3.5. Citation classification

There are parallels between the typing of links and the typing of citations. Since citation databases first became available authors have proposed systems for categorising authors' motives and/or the rhetorical role citations play, e.g. Lipetz (1965), Weinstock (1971), Murugesan and Moravcsik (1978), Duncan et al. (1981) and Garzone and Mercer (2000). While these schemes are varied they share common elements, such as corroborating/affirmative, negational/correcting, methodology, background/assumed knowledge, which we recognize also among the relational types of our own ontology and Gil's. We are convinced that citation indexes would be greatly improved by this kind of typing and are investigating its application in other projects. However to be economic it must be automated and this is a substantial challenge for natural language researchers. Much of the automatic classification of citations carried out to date has been aimed at document summarisation and argumentative zoning (finding the parts of papers that play different roles) rather than directly at citation classification, e.g. Nanba and Okumura (1999) and Teufel and Moens (2000). It is an interesting observation that these authors employ very basic categorisation schemes of just three or four key types. Nanba and Okumura have

> Type B—the references to base on other researchers' theories or methods, Type C—the references to compare with related works or to point out their problems, Type O—the references other than types B and C,

whereas Teufel and Moens have Background, Other work, Weakness/Contrast and Own contribution. By contrast, Mercer et al. are investigating textual cues to mark up a two tier system with 34 base types divided between 10 upper categories (Mercer and Di Marco, 2004).

We are convinced that citation indexes would be greatly improved by this kind of typing. Although at present, it is fair to say that these techniques are still in a relatively early state of development. Apart from providing an automated technique to apply to a specific document corpus (once properly trained), the key difference to our approach is that the granularity of our work is the *claim*, as opposed to the *document*. The complementarity between the two approaches hold potential, however, and Sereno et al. (2005) have reported the evaluation of a prototype system which applies Teufel and Moens' (2002) argumentative zoning and other information extraction techniques to more actively support the task of detecting and annotating, potentially significant claims in documents.

4. The tools

Three prototype tools were used in the case study reported in Section 5 of this paper. The first is ClaiMapper, a sketching tool that supports users in making sense of the claims in papers. When the user is satisfied with part or entire claim network produced in ClaiMapper it can be imported into the second tool ClaiMaker. This is a digital library server that connects claims via hyperlinks to the documents they describe and provides search services to help users explore large claim networks. The third tool is ClaimFinder. This is an alternative interface to ClaimMaker, designed for use by novices, which contains the simpler search functions presented more accessibly.

4.1. ClaiMapper argument sketching tool

The first prototype for building claim networks was a form-filling interface which can be accessed directly through the ClaiMaker server. This interface has one form to create Concepts, another to create Sets and a variety of forms for creating different kinds of Claims. This was a quick route to allow the research team to start populating the database with claims in order to put the ontology through its paces and create a collection for testing services, but it did not provide much sensemaking support to modellers. For instance, while the project team did become adept at choosing from among the many options

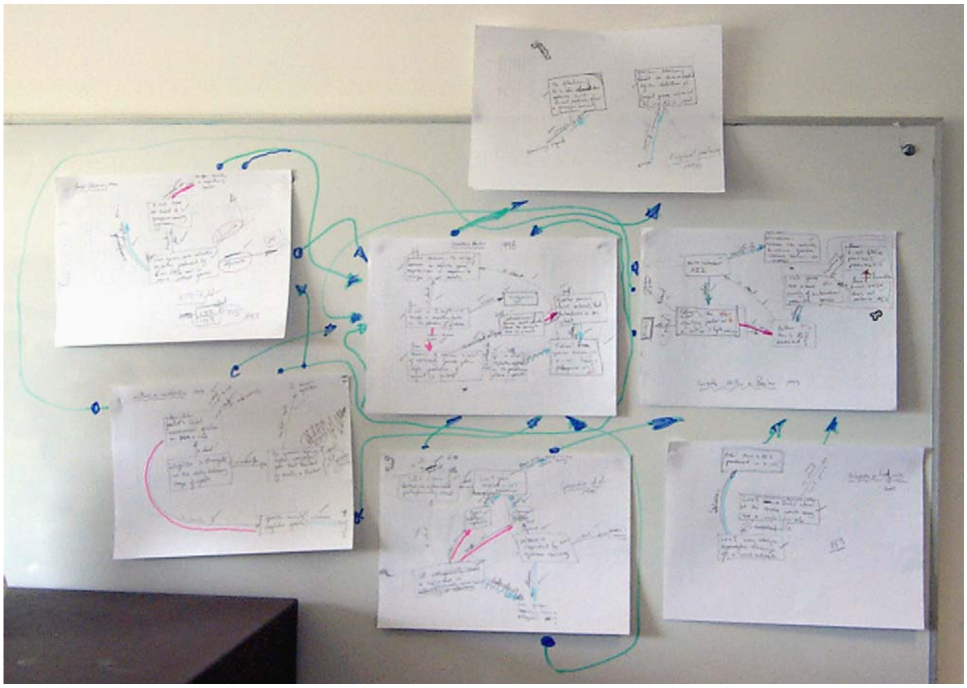


Fig. 4. Example of sensemaking using pen and paper sketches of claim networks.

on the interface, they reported difficulty holding a gestalt view of the model in their heads as they went through the dissociated steps of building Concepts and Sets, then assembling them into Claims. It was clear that some radical changes were needed to the interface if it was to support the cognitive processes involved in creating claim networks.

In order to overcome the problems of holding complex models in memory, the project team sometimes found themselves resorting to pen and paper for sketching out drafts of argument maps. An extreme example showing a paper-based concept network describing several papers is shown in Fig. 4: each sheet of paper has a sketch of the claims in one paper, the arrows drawn between sheets on the white board represent claims which use concepts from two different papers. This sketching stage was adopted in part because the form-filling interface had no correction facility. It prevented users from deleting or modifying a concept which might meanwhile have been included in someone else's model. However it was mainly driven by a desire to refine the user's own interpretation before committing it to the knowledge base. In the terms of Green's cognitive dimensions analysis the interface was "enforcing premature structure" (Green, 1990; Shipman and Marshall, 1999), by making the users commit a structure before they were comfortable that they had made sense of what they were reading.

It was clear that a new interface should offer better support for this sketching stage, which enables the refinement part of the sensemaking process. It therefore had to assist the process of sketching draft maps and reviewing new structures in context before committing them to the knowledge base. This was implemented by modifying Compendium, a hypertext visual modelling tool.[4] The result was a desktop sketching tool called ClaiMapper in which a small number of icon conventions are used to produce visualisations of concepts and the connections between them.

The ClaiMapper conventions are illustrated in Fig. 5. The right pane is an open *document* that contains two *concepts* (represented by the orb icon) and a *set*, which contains two other *concepts* (represented by the bullet list icon with the subscript 2 indicating the number of *concepts* contained in the *set*). These are linked to form two Claim triples. A Claim triple comprises two objects joined by a directed, labelled link. We refer to the objects conventionally as the left- and right-hand objects, the left hand being the place the link comes from and the right hand the place it links to. Of the claims in Fig. 5, one has on the left hand the set and on the right hand one of the concepts linked by the relation *is analogous to*. The second claim has on the left hand the concept labelled *One Claim can contain another* and on the right hand the whole of the first Claim triple (represented by the *is similar to* link pointing to the centre of the *is analogous to* link). Using ClaiMapper in this way we can clearly visualize the nesting described in Section 2.1.

The structures in the right *document* of Fig. 5 are structures that can be uploaded to the ClaiMaker server and analysed. However the ClaiMapper tool does not restrict the user from making other kinds of informal structure that are helpful to the organisation and refine-

[4]Compendium semantic concept mapping tool: www.Compendium-Institute.org.

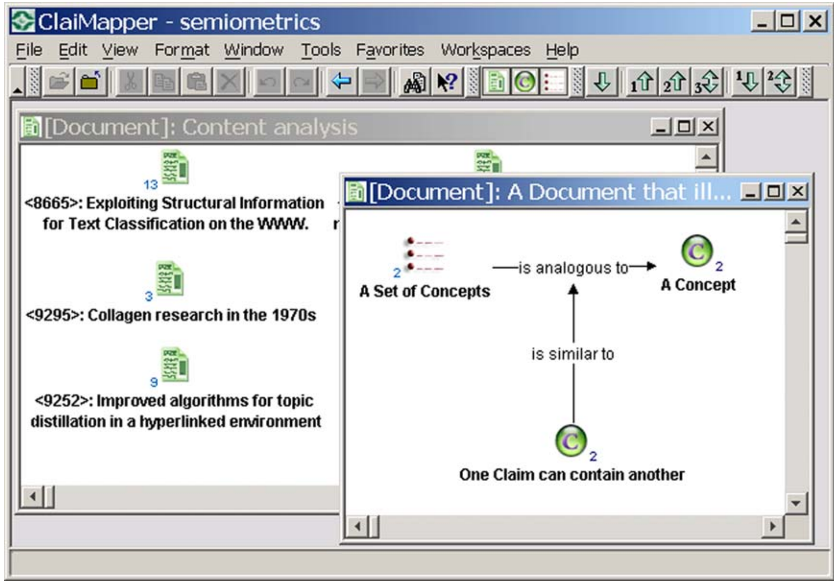


Fig. 5. Modelling claims in ClaiMapper. In the left pane is a collection of documents (the digit on each icon indicates how many concepts are annotated on it). Double-clicking one of these opens a new window, e.g. on the right, showing two Concepts, a Set and two Claim triples, one of which is the right hand of the other.

ment of their ideas. For example, the left window in Fig. 5 is being used as if it were a folder to hold a collection of *documents* on the same topic.

While the user works in the ClaiMapper environment they can edit the structures they build at will. However in order to access the interpretational services provided by ClaiMaker they will eventually have to decide that their interpretation of a particular document is "good enough" to be uploaded to a private or shared database. At this point the structures that have been uploaded are synchronized with the server by being given unique IDs (see Fig. 15b, IDs are numbers in angle brackets). Subsequently ClaiMapper prevents editing of those structures, for example, changing a label, which would cause a mismatch with the server version. The user is still free to link to them, in which case the additional claims can be uploaded to the server. Structures can also be deleted from the ClaiMapper version, since the server version does not require users to have a full copy of the server model on their local machine it will not detect an error. The user can also download their own and other modellers' structures from the server for use in ClaiMapper.

4.2. ClaiMaker server software

The ClaiMaker server combines a number of roles. It supports model building, through the original form-based interface and through model upload via XML files exported by ClaiMapper. It also provides a range of search, visualisation and discovery services. Implementing ClaiMaker as a server application has a number of practical advantages. In the development stage it facilitated getting new versions to early adopters; changes made to the server are available to users without them having to regularly update their local software. It also gave the project team access to the models people built allowing us to assess the modeling scheme and identify difficulties. Uploading claims once to a server is a much easier way for a distributed group of collaborators to share their annotations than circulating files of annotations which each member of the group must upload individually to view them. Finally, to support ClaiMaker's role as a digital library system, the choice of a server, in which links can be made to digital resources via URLs, is obvious.

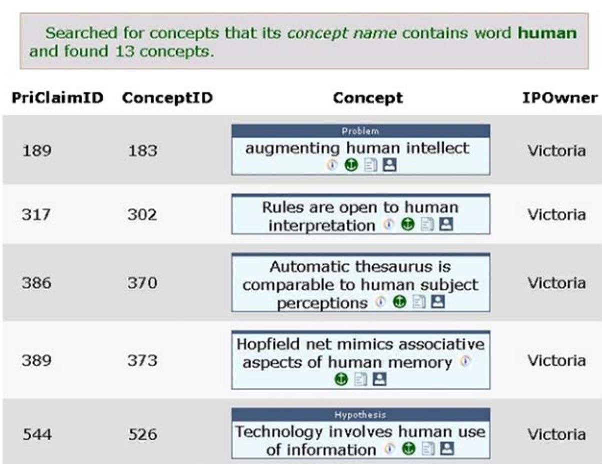
The data (*Concepts*, *Sets*, *Claims*, bibliographic metadata, etc.) are stored in a MySQL database which underlies all the functions. A mirror of the database in RDF was also maintained at one stage on a Lisp server. This allowed us to experiment with services that exploited the structure in the link ontology to a greater extent. We were able to develop some interesting services using this technology, e.g., *lineage*, which will be discussed further below. However, we found that the technical difficulties of maintaining two intercommunicating servers were too great for us (and probably for potential users as well). Consequently we found ways to recast these services as complex SQL statements that replicate most of the functionality of the RDF-based search.

We will not describe the model building functions of ClaiMaker, since the form-based interfaces it uses have been largely superseded by the sketching technology used in ClaiMapper. We will concentrate instead on the search, visualisation and discovery functions. Later we will demonstrate how some of these functions were used to analyse the claim network produced using ClaiMapper.

4.2.1. Search

ClaiMaker has a basic set of services that allow the user to find *documents*, *concepts*, *sets* and *links* by searching for title text, author, creator, creation date, etc. As a first step these give tabular listings of the type shown in Figs. 6 and 7. From these listings further information about the objects returned can be found using the icons located with them. These open supplementary information boxes that overlay the main listing. Working from left to right on the *concepts* in Fig. 6 the "*i*" icon will bring up further information about the concept itself (who created it, when, any further details input by the creator, etc.). The "*anchor*" icon sets this concept as the focal node in a *Neighbourhood* search (described shortly). The document node gives the bibliographic details of the document that backs the concept up and a hyperlink to the original. The final "*person*" icon gives details of the person who created the object.

The listing for links (see Fig. 7) shows for each claim the two objects that are joined in the first and third columns with an arrow icon between them. The same icon set is used to access additional information and the colour of the arrow depends upon the properties of the link so that red links represent links with negative polarity and green links have strong weight and positive polarity (the rest are gold). Referring back to Table 2, "*improves on*" is shown with a green arrow, "*is different to*" with a red arrow and less strongly negative/positive relations such as "*is evidence for*", "*addresses*" and "*is about*" with gold arrows.



Searched for concepts that its *concept name* contains word **human** and found 13 concepts.

PriClaimID	ConceptID	Concept	IPOwner
189	183	Problem: augmenting human intellect	Victoria
317	302	Rules are open to human interpretation	Victoria
386	370	Automatic thesaurus is comparable to human subject perceptions	Victoria
389	373	Hopfield net mimics associative aspects of human memory	Victoria
544	526	Hypothesis: Technology involves human use of information	Victoria

Fig. 6. Part of the listing of results from a concept search for the string "human".

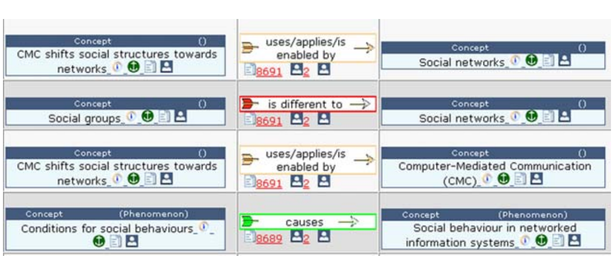


Fig. 7. Part of the results listing for a search for links.

4.2.2. Visualisation

In addition to the tabular layout search results which contain claims can be viewed using interactive views generated using the TouchGraph[5] visualisation package. For example Fig. 8 is a TouchGraph rendering of search results. The visualisation can be explored by the user via the *locality*, *zoom* and *rotate* functions, or filtered by link type using the menu shown. Further information on nodes in the display can be accessed by hovering over a node and selecting "*details*", as shown.

4.2.3. Discovery

Developing discovery services has been a core activity within the Scholarly Ontologies project. Traditional information retrieval systems use term-based search and search via citations. Term-based search handles documents as isolated entities defined by the words in them. Citations in a document do give an indication of the links between documents but there are many motives for citing and a reference list gives no indication of authors' intentions in referring to other work. We generally cannot even tell if a paper is referenced because the authors support it or are diametrically opposed to it, although interesting research is being done to improve this situation (see Section 3).

In this section, we describe four of the discovery services that we have developed. The examples of discovering the neighbourhood and discovering chains, demonstrate services to assist the user in exploring and navigating the topologies of argument maps. The examples of discovering disagreement and discovering lineage demonstrate how the explicit connections embedded in the discourse ontology can be used to build services that assist the user in answering common research questions, e.g. "Where did this idea come from?" A typical discovery service comprises a search of the claims network followed by a presentation of results (which may be a visualisation) tailored to the particular question.

4.2.3.1. Discovering the neighbourhood. The *Neighbourhood* search, which can be reached from tabular results listings via the *anchor* icon gives answers to the question "What is directly related to this?" It allows the user to examine all the claims made with one or more chosen object/s on either the left- or the right-hand side. The focal concepts can be searched for using a keyword search, or a search for all the concepts in a particular paper, or by picking up an *anchor* icon from a previous results listing, or by picking up an *anchor* icon from one of the left-or right-hand column headings in the neighbourhood table which selects all the concepts in that column as focal concepts. The focal concepts are listed in the central column of the *neighbourhood* listing (see Fig. 9). Because it embeds the

[5]TouchGraph LLC, www.touchgraph.com.

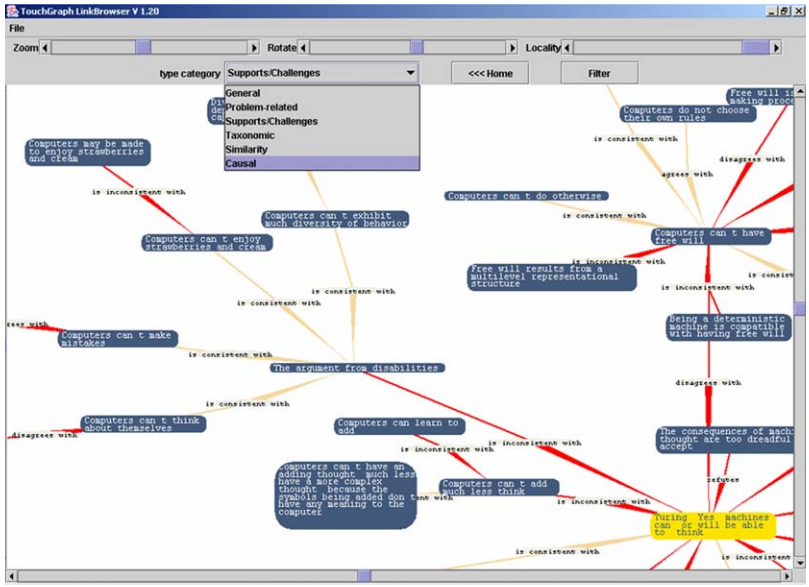


Fig. 8. TouchGraph interactive visualization of branches in a claim network modelling literature from the Turing Debate on machine intelligence, emanating from the root node (lower right) "Turing: Yes, machines can or will be able to think". (We gratefully acknowledge Robert Horn's paper maps as the source for this example: www.macrovu.com/CCTGeneralInfo.html.)

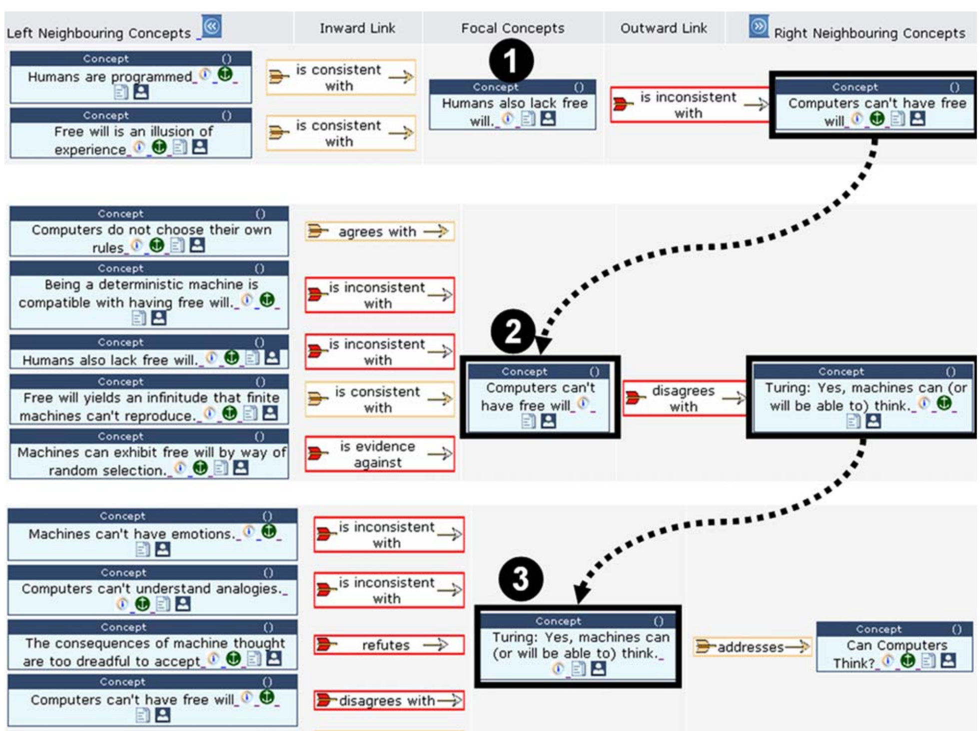


Fig. 9. Navigating the 'Neighbourhood' around a concept. Clicking on a concept makes it the central node and displays the incoming and outgoing links one step away. (We gratefully acknowledge Robert Horn's paper maps as the source for this example: www.macrovu.com/CCTGeneralInfo.html.)

anchor icon in its own results listing (see Fig. 9), the *neighbourhood* service allows users to step through complex networks following routes that are interesting to them.

4.2.3.2. *Discovering chains.* The *link-tracking* service has similarities to *neighbourhood* but allows the user to specify more parameters, for example the length of chains to be found and their direction, by filling in slots in a simple form. Fig. 10 presents an example of the output of this service in TouchGraph format for a search for chains of one link out from any Concept on the left hand of a claim triple containing the string "CiteSeer".

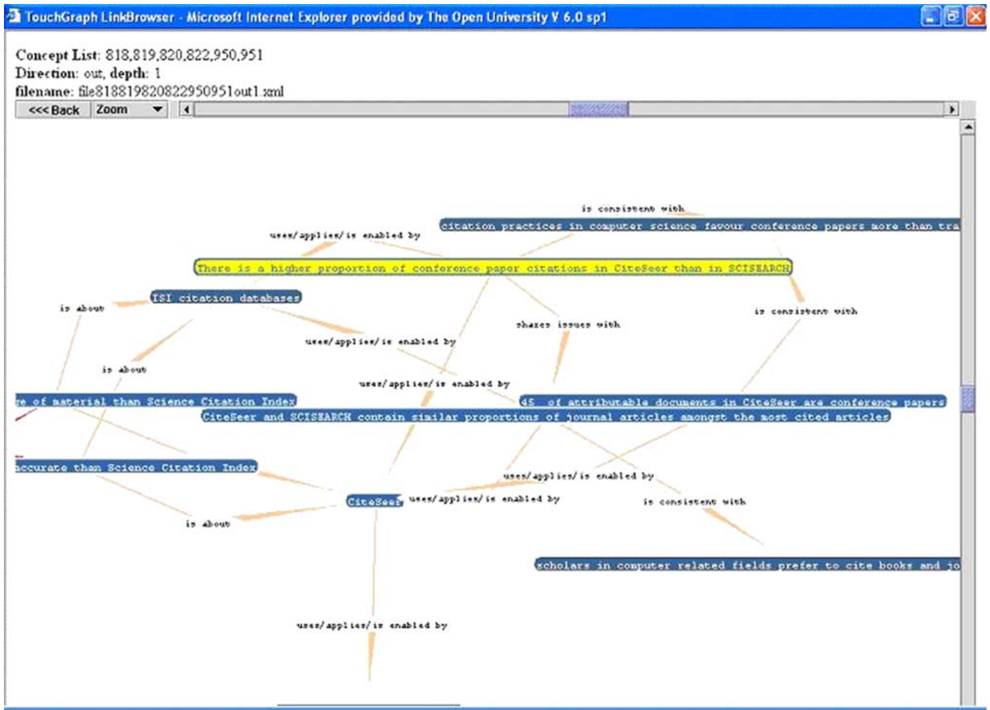


Fig. 10. TouchGraph presentation of results from a link-tracking search.

4.2.3.3. Discovering disagreement. Consider a common question that many researchers bring to a literature: "*What arguments are there against this paper?*" Despite the centrality of the notion of disagreement in the assessment of evidence, there is not even a language in which to articulate such a query to a digital library. With our ontology modelling the world of scholarly discourse, we can begin to express the basic idea that *researchers disagree*; it is the idea embedded in the property negative polarity.

How can we operationalize such a query? First, we are looking for *arguments against*, which map on to the ontology as relations of any type with negative polarity. At a trivial level, *this paper* corresponds to the currently selected document in ClaiMaker. But more substantively, *this paper* refers to the *claims* that researchers have made about the document, specifically, the *concepts* linked to it. Moreover, we can extend this to *related concepts*, using the following definition: *the extended set of concepts linked by a positive relation to/from the document's immediate concepts*, i.e., discovering chains of disagreement or agreement.

For the given document, this discovery service does the following:

1. Finds the concepts associated with that paper.
2. Extends the set of concepts by adding positively linked concepts from other papers.
3. Finds concepts that link to these with negative relations.
4. Returns the concepts from step 3 as concepts against the extended concept set.

This approach has dangers. It does not follow that if A is in agreement with B and B is in disagreement with C then A must be in disagreement with C also. However it should be remembered that this is a search service. It is up to the user to judge whether the claims returned are valid.

Typical results are presented in Fig. 11. Note the two numbers to the right of the claim that *disagrees with* one of the related issues in the query. The first (8621) is a hyperlink to the metadata of the paper that provides the backing for the claim, which includes a URL to the paper itself. The second (2) is a link to the personal details of the modeller who made the claim; this allows the user to make a judgement about the credentials of the claim: can it be trusted?

4.2.3.4. Discovering lineage. A common activity in research is clarifying where a particular idea came from and what other ideas influenced its development. We call this the *lineage* behind an idea. *Lineage* is the notion that ideas build on each other and has an inverse, the *descendant*, which is the notion that ideas are spawned by a particular seminal notion. Where the paths have become increasingly indirect over time or been confused, uncovering unexpected or surprising lineage is a major scholarly contribution. We have a more modest goal to start with in ClaiMaker: to provide a tool pick out from the 'spaghetti' of claims, candidate streams of ideas that conceptually appear to be building on each other.

In practice, our *lineage* tool tracks back (semantically, not in time) from a concept to see how it evolved, whereas

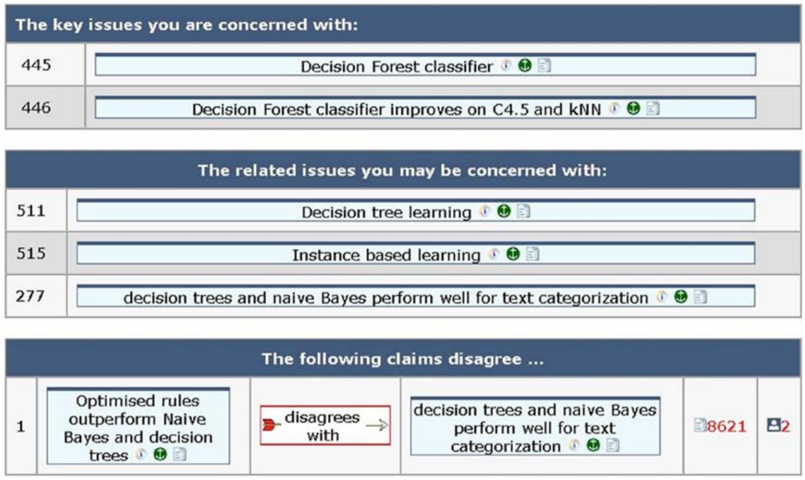


Fig. 11. Arguments that contrast with the concepts in a research paper (Chen, H., Ho, T., 2000. Evaluation of decision forests on text categorization. In: Proceedings of the Seventh SPIE Conference on Document Recognition and Retrieval).

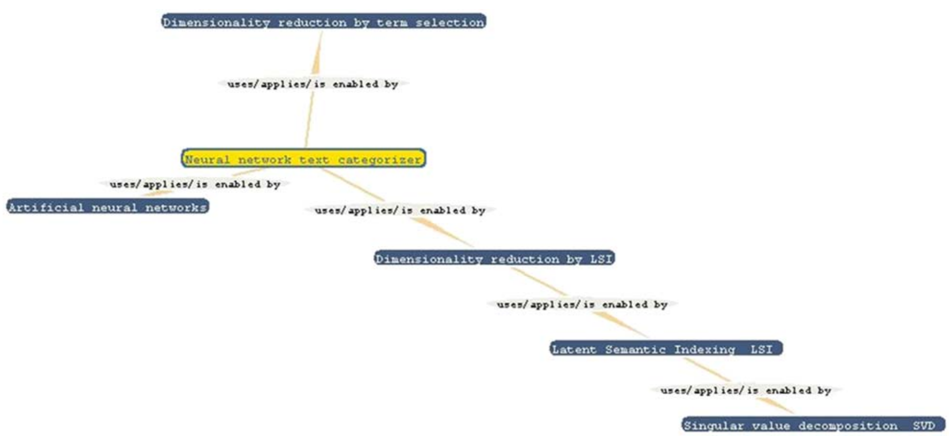


Fig. 12. Output of a *lineage* search from the node "neural network text categorizer".

the *descendants* tool tracks forward from a concept to see what new ideas evolved from it. Since *descendants* are the inverse of *lineage* (and are implemented as its literal inverse) we will only discuss *lineage*. A *lineage* can be conceptualized as a path in which the links suggest development or improvement (Fig. 12).

The constraints are chosen to reflect the ideas of improvement and development. The set of permitted link types comprises the two general links *uses/applies/is enabled by* or *improves on* and links of type *similarity* and positive polarity. The *improves on* link type is included to reflect the notion of improvement, while *uses/applies/is enabled by* has the weaker implication of development. The *similarity* links are included because if a new concept is like a second that *improves on* a third concept, then the new concept is likely to be an improvement on the third concept as well. The problem of finding *lineage* in ClaiMaker can then be formulated as a path-matching problem, a well-known problem in graph theory, which searches for paths (sequences) of links that follow a specific pattern. The first prototype of the lineage service used an RDF representation of the argument maps and the Ivanhoe path matcher embedded in the Wilbur RDF Parser (Lassila, 2001). This approach is described in our other papers: Buckingham Shum et al. (2003) and Uren et al. (2003). Due to operational difficulties with supporting a Lisp server in parallel with the ClaiMaker server this approach was later dropped and we adopted a slightly weaker approach to *lineage* based on a chain search with the links going away from the home node, pruned using constraints based on the link ontology. The *descendants* algorithm is the same except that the links in the chain are directed towards the home node.

The procedure is as follows:

(1) The user inputs a home node H and a number of steps N, the maximum length of lineage they wish to search through.
(2) Find all links in the direction away from H (H is the left-hand side of the triple) that meet the set of

constraints (type similarity and positive polarity, *improves on*, *uses/applies/is enabled by*).
(3) Define the output set H' containing all the terminal nodes of the paths found in step 2.
(4) Eliminate from H' any nodes that have been encountered before in this search.
(5) For each node in H' repeat steps 2 and 3 and build a new set of terminal nodes to search from.
(6) Repeat steps 3 and 4 until a total of N cycles have been completed.

Two things were lost in the MySQL implementation, compared to the first RDF-based prototype. The first is that we can no longer handle similarity links as symmetric links, i.e. we can no longer ignore their direction. The second is that, because MySQL has poor facilities for recursion by comparison with Lisp, the user must now set the maximum number of links they wish to see in a chain.

4.3. ClaimFinder search interface

ClaimFinder was written as a novice interface for the search and discovery services provided by ClaiMaker. The functionality of the services is the same and they access the same databases, only the presentation is different. Fig. 13 illustrates the "*Find*" screen. This is the first screen a user comes to on entering ClaimFinder. The user can enter keywords, which will be searched against the text of Concepts. He can also select using radio buttons whether the output will be represented as a table, which will be a *neighbourhood* table as illustrated in Fig. 9, or a graph which will be a TouchGraph visualization like Fig. 10. All subsequent screens retain this basic "*Find*" search entry box in the banner.

The other tabs on the find screen allow the user to access the other services.

- Discover gives access to *Contrast* and *Agree* and *Lineage* and *Descendants.*
- Advanced allows the user to search the database for *Article Title*, *Article IDs*, *Concept creator*, *Keywords* in concepts (i.e. the same as *Find*), *Concept IDs* and *Concepts* added in the last X number of days, where the user specifies X.
- ClaiMaker takes the user to the ClaiMaker "expert" interface.

Fig. 13. ClaimFinder interface "*Find*" screen.

5. Empirical evaluation: creating and reusing a multi-disciplinary model

The study had two phases, to address both the 'writing' and 'reading' of these new forms of scholarly artefact. The first was a modelling phase in which a Claim Network was built and a short review of the topic was written. The second phase evaluated the affordances of these two artefacts for communicating to users other than their creator via a factual questionnaire.

A real research task was required to test the modelling tools. We chose to examine a multi-disciplinary domain at

Table 3
Summary of topics covered by the case study review

Link-based analysis methods
Scientometrics, the study of scientific research literature using citation data, has been used to study the development of ideas and identify emerging topics for some years(White and McCain, 1989; van Raan, 1997). The process generally involves the selection of a body of citation data in a field of interest, followed by computation to identify structures of interest, which are then analysed by an expert to interpret what the structures mean in terms of the development of the field.
Scientometrics overlaps with the study of *social networks*, which models nets of relationships between people (Pool and Kochen, 1978/79), via co-authorship studies. In this case, the relationship link is made if two authors have published together and may be weighted according to the number of co-authored publications.
The extension of scientometric practices to analyse the World Wide Web, sometimes called *Webometrics*, is being actively explored, although not without caveats concerning the signification of hyperlinks (Cronin, 2001). For scholarly hyperlinks, studies have shown that researchers motives for hyperlinking are closely related to their existing citation behaviour (Kim, 2000). Therefore it may be reasonable to assume that scholarly hyperlinks are suitable for scientometric study.
This view is reinforced by the successful development of *Web ranking algorithms* that exploit information about the links between Web pages, e.g., PageRank (Page et al., 1998) and HITS (Kleinberg, 1999). Using link information as part of the ranking strategy is considered advantageous because links from domains other than a page's home domain represent some kind of human endorsement of the content.

the intersection between scientometrics, co-authorship studies and Web link analysis methods. A summary of the topics is given in Table 3.

A number of motives influenced this choice:

- The field was characterized by the common theme of link analysis.
- Citation studies and co-authorship studies are well-established methods of literature analysis.
- Web-based link analysis is currently a hot topic in information retrieval.
- The topic seemed likely to inform our own research on claim networks.

5.1. Sketching and refining interpretation with ClaiMapper

In the first phase of the evaluation we used ClaiMapper and ClaiMaker to build a claim network. The first author undertook the review and initial modelling of the literature in ClaiMapper. By the end of this process some 290+ nodes, 340+ links and 64 document nodes had been created. Ideally, we would have like to have had the Claim Network built by a neutral third party. However, since the person who built the review had to be familiar with the tools and willing to commit several weeks of working time, we were constrained to select one of ourselves.

It was observed that the overall sensemaking process with ClaiMapper fell into three cycles, which overlapped and followed in sequence, but with backtracking (see Fig. 14). The initial cycle is the Gather–Read–Categorize cycle, in which a collection of potentially useful material was obtained, scanned and roughly sorted into topics. The middle cycle is the Read–Model–Categorize cycle. Here the papers were studied in more detail, the arguments were mapped and refinements were made to the categories used. The final cycle was the Model–Reflect–Write cycle, in which the claim networks were used to draft summaries on each of the topics. Backtracking occurred, e.g., when writing and reflecting opened up new questions which required more documents to be gathered.

ClaiMapper provided a space in which to sketch rough ideas, then refine them. It was found that, in this case

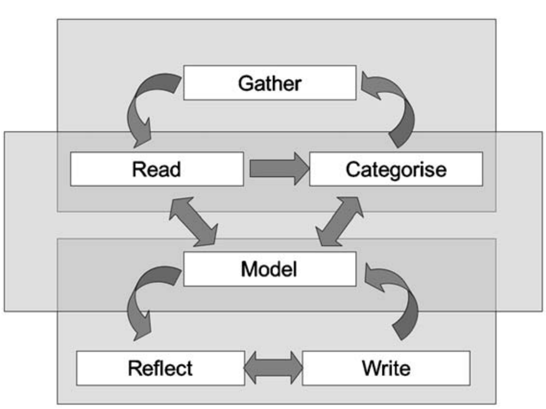


Fig. 14. Review Processes using ClaiMapper.

study, the degree of order in the models seemed to increase over time. For example, in Fig. 15a the first screen shot is taken from a backup of the ClaiMapper database taken at a relatively early stage, roughly at the end of the first Gather–Read–Categorize cycle, whereas the second screenshot (Fig. 15b) was taken after several Read–Categorize–-Model cycles, while writing was in progress.

In Fig. 15a, the *Home Window*: is being used as a scribble pad. *Concepts* of interest are dotted about and linked to each other and to *documents*, some represent research articles while others are being used as containers to organize material, rather like a folder in a hierarchical file system. One of these containers has been opened; and contains one *document* and a series of unconnected *concepts*. At this point, the structure has some of the aspects of an argument model, related ideas have been joined up at the upper level, but it is largely mnemonic: a sketch of ideas that arose from the initial scan and deserve further investigation.

By the time the screenshot in Fig. 15b was taken, the structure of the models had become more organized. The *Home Window* now contains just eight *documents*, each of which is acting as a container for *documents* on a particular topic. The containers have been organized into a shallow hierarchy with "bibliometrics" at the top. One of these container *documents* has been opened. It contains an unconnected list of *documents* each of which represents an actual article. The right hand small pane shows the argument model for one of these articles. This is expressed using the constraints of the ontology described above and is a machine-interpretable structure that could be uploaded to ClaiMaker as a representation of this document.

We can see in this process that ClaiMapper was able to support the refinement aspect of sensemaking. It appears that as the modeller learnt more about the topic and became more committed to her interpretation of the data, a crystallisation process occurred in which the models became more organized and clear categories emerged. It is a classic example of the move from rough sketches to coherent argument, a phenomenon reported both in empirical studies of designers using argumentation-based design rationale and empirical studies of the use of computer-supported writing tools, reviewed in (Buckingham Shum and Hammond, 1994).

The use of *documents* as containers in which to subdivide papers by category in this case study may have arisen from the multi-disciplinary nature of the task. This is an interesting example of affordances of the system emerging that were not designed in as functions. One reason that *documents* could stand in as containers is that ClaiMapper permits transclusion,[6] so papers which bridged categories

[6]In Hypertext research, 'transclusion' is a term invented by Ted Nelson for the republishing of the same content in multiple contexts, such that the system treats the material correctly and the end-user is aware of the reuse. In ClaiMapper (and Compendium on which it is based) transclusion manifests as nodes which can be edited directly from any of their contexts, which display where they are transcluded and support quick navigation between these contexts. See Selvin and Buckingham Shum (2005) for an

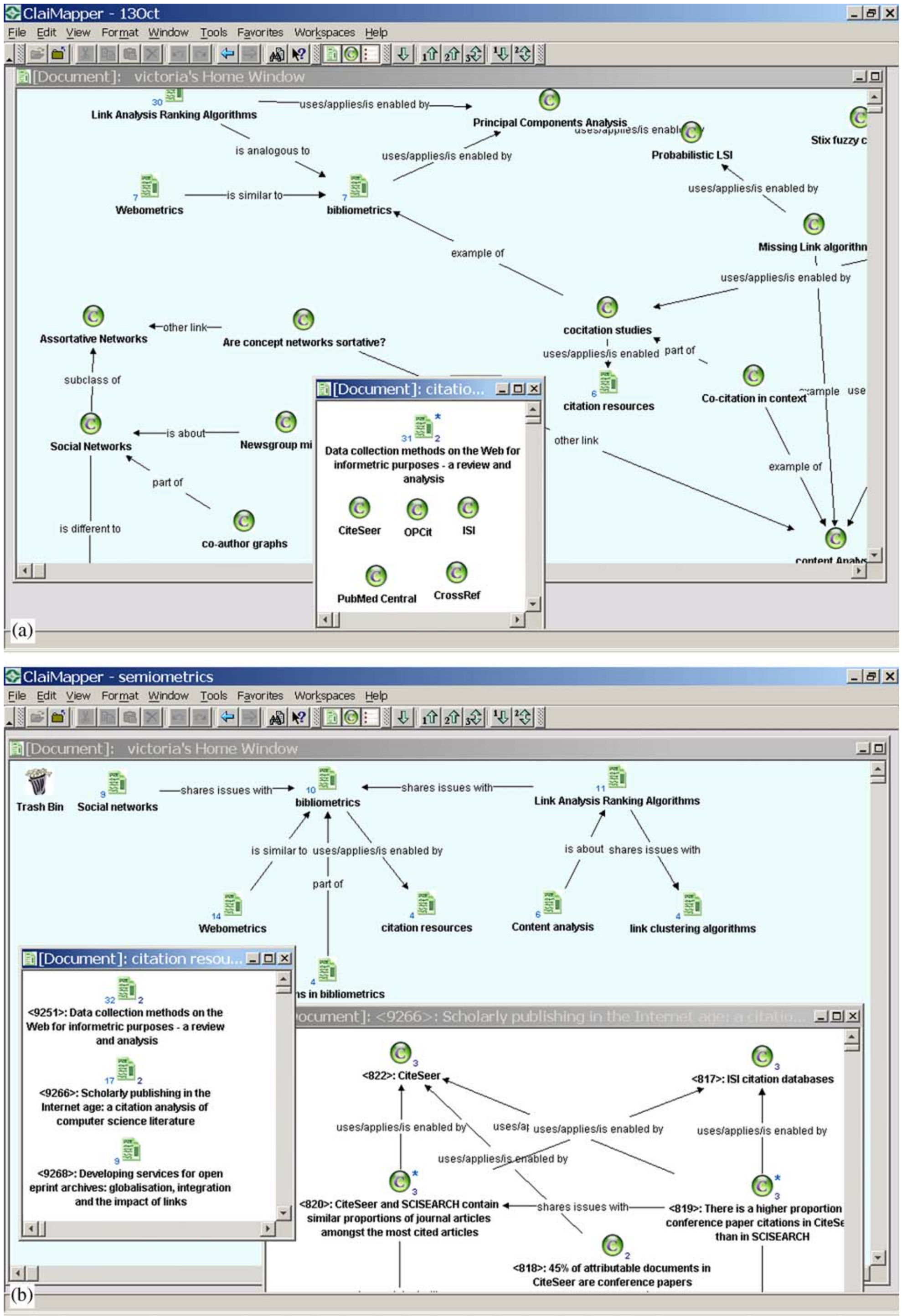


Fig. 15. (a) Screenshot of ClaiMapper at an early stage of the modelling process and (b) Screenshot of ClaiMapper at a late stage of the modelling process showing the crystallisation of interpretation over time.

could be copied into in more than one container and edits made in either place would appear automatically in the other.

(*footnote continued*)
account of how this can assist knowledge management and sensemaking in long-term research projects.

The hierarchy shown in Fig. 15b was used to provide the sectioning for the first draft of the review. This sectioning was changed later because it did not give sufficient emphasis to the cross disciplinary threads that made the review interesting. However it provided the structure for an initial "divide and conquer" step in the writing process, in which topic summaries could be produced simply by looking at the documents in a particular "folder".

Clearly we cannot draw general conclusions from observations of one individual's use of a tool, but it concurs with the experience of other team members and more broadly with the way in which we know learners gradually elaborate their understanding of a domain through concept splitting, merging and clustering. ClaiMapper integrated into the natural review process smoothly, adding value in the early stages by providing a scribble pad on which initial observations could be sketched and in the later stages it provided concept manipulation tools for ordering observations and providing notes on them in a way that supported writing and reflection on the material.

5.2. Comparative evaluation

The second phase of the evaluation comprised a comparative analysis of the two artefacts produced in phase one, namely the Claim network and the traditional written review. We wanted to determine whether the Claim network, in combination with ClaiMaker and ClaimFinder, could be used to communicate similar information to that in the written review. This is the first step in validating Claim networks as a collaborative tool—demonstrating that one person can understand another's model. A factual questionnaire was used as the instrument of the study. In particular, we wanted to look at the overall quality of the students' answers, how they handled individual questions and which features of the search tools the claim network group used to find answers.

There are some confounding factors in the design of this experiment. As we have already commented, ideally the artefacts should have been built by a neutral third party. However this was infeasible. Also the quality of both the claim network and the review were likely to influence the quality of the answers that the students gave. However, the question we were seeking to answer was not very complex. It was simply whether claim networks could communicate information to people other than the creator. The review group give a point of comparison but differences between the media, the history of their preparation (the review was written after and based on the claim network and the questionnaire was written last) and the fact that it was impossible to make the content of the two artefacts identical force us to be cautious about the significance of any differences in the performance of the two groups.

The participants in the study were six research students studying in KMi. None of them had prior, in-depth knowledge of the topics in the literature selected for the study. Half the group was engaged in research related to discourse mapping and literature analysis. These three were all familiar with the ScholOnto discourse ontology and the ideas underlying claim networks but were not particularly familiar with the tools. These students were assigned to the *Claim Network* group. The remaining three students were assigned to the *Written Review* group and worked with a document of about 2300 words in length. It was not considered detrimental to the study to use the students with knowledge of the principles of claim networks to use the tools, since we wished to investigate a scenario in which the basic ideas and instrumental operations were known (just as members of the Written Review group were familiar with reading, pens and paper). Even after deliberately selecting students with some knowledge of claim networks the bias of experience of the medium used still favours the Written Review group.

Table 4
Questionnaire

Evaluation Questionnaire
1. What are the disadvantages of using a Web crawler to collect data?
2. Name four algorithms for ranking Web pages.
3. Select one which you consider particularly important and explain why.
4. What is scientometrics?
5. What does van Raan consider to be the sub-tasks of scientometrics?
6. What problems arise when applying scientometric methods to Web data?
7. Name three properties you would expect to see in social networks.
8. What advantages and disadvantages does CiteSeer have compared to the ISI citation databases?
9. Give the titles of two papers which report on combining information from Web pages with link analysis algorithms.
10. What unifying notion is common to scientometrics, social networks studies and link ranking algorithms?
11. If you were to undertake a small research project in this field what part of it would you choose to tackle? Please explain your choice.

A questionnaire was written which could, in principle, be answered using the information provided by either artefact (see Table 4). A testing station was set up with the Camtasia[7] screen capture tool to record the participants' interactions with the tools and their verbal comments. While the verbal comments were not used heavily in the analysis reported here, the comments of the claim network group were a valuable source of qualitiative data to understand why they were pursuing particular strategies and to identify design flaws and bugs in the interface. Participants were accompanied by an observer who could assist with any general queries they had about the exercise and who also provided someone to "think aloud" to. The questionnaire was presented on screen to facilitate timing how long it took to answer each question. Camtasia recorded participants as they added or edited their answers in this online version and the time taken on each question was estimated from verbal comments and time spent inputting answers. The Claim Network group was given a Microsoft Internet Explorer Web browser with links set up to both ClaiMaker and ClaimFinder. The Written Review group had an open Microsoft Word document containing the review, plus a hard copy version since many people prefer to read on paper.

[7]Techsmith Corp., Camtasia Studio: http://www.techsmith.com/products/studio/default.asp.

5.2.1. The questionnaire and correctness scoring system

Subjects were required to answer the questions listed in Table 4. The aim of the questionnaire design was to cover a spread of the topics in the domain and to give questions that ranged in complexity from extracting facts to demonstrating some understanding of the topics. Taking some examples: question 2 concerns the topic of link ranking only and simply requires the participant to identify some names of algorithms, question 6 concerns both scientometrics and the Web and expects the participant to be able to identify issues that are problems and question 10 concerns all the topics and requires the user to form an overview of them. The openness and complexity of the questions tends to increase towards the end of the questionnaire to give students the opportunity to acquire some knowledge of the domain.

To assess the correctness of the results, we constructed a "gold standard" set of answers by merging the answers of the three individuals who tested the questionnaire prior to the study. These were the first author, who had access to the review and the claim network, the second author, who only used the claim network and an experienced research student who was not part of the research team, who only used the review. A marking scheme was devised and the third author, who was not involved in any other part of the evaluation, marked the answer scripts.

The scoring system was weighted equally across the questions; it allotted a maximum of two marks per answer, giving a maximum score of 22. For factual questions, a list was supplied of all the items listed by the gold standard group in answers and a suggested mark was allotted for each item up to a maximum of 2. Question 2, for example, had half a mark per algorithm up to a maximum of 2 marks. The exception was question 7, which asks for precisely three properties In this case a bonus half-mark was awarded if the participant gave exactly three. For the more open questions the marker had to use some discretion and judge how well argued the answer was and whether it included reference to any of the issues raised by the testers.

We also recorded the times participants took to answer individual questions to determine their relative difficulty. The decision of a participant to move from one question to another was a cognitive process which could not always be timed with precision. The timings were therefore measured in minutes and rounded up. Minutes gave sufficient accuracy to get a feel for the relative difficulty of questions, which was our main aim.

5.2.2. Results

Our analysis of the data in the Camtasia movies and the students' answers to questions covered several aspects. Correctness and actual time taken to answer questions gave us an indication of the comparable difficulty of the two tasks (review or claim network). The per question analysis of relative time taken and answers given helped us understand whether there were affordances of the two artefacts which were advantageous in particular situations.

Table 5
Correctness and time spent on the exercise by each participant

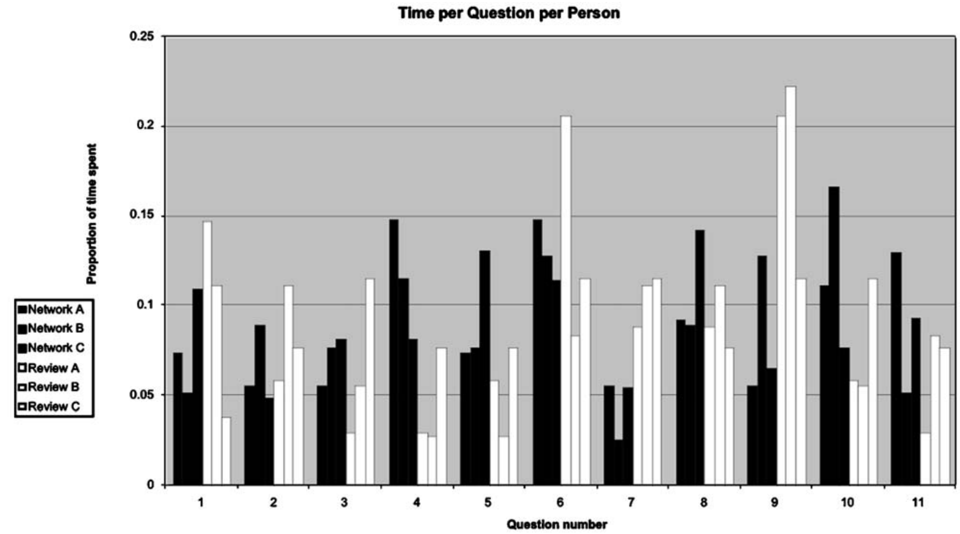
Task participant	Correctness (max. 22)	Approx. time in minutes
Network A	9.5	54
Network B	13.5	78
Network C	15.5	183
Mean network	*12.8*	*105*
Review A	11.5	56
Review B	14.0	36
Review C	17.0	38
Mean review	*14.2*	*43*

We gave special attention to the questions which required the participants to interpret the material they were given. An analysis of how the participants in the Claim Network group used the available interfaces informed us about which services were found to be most helpful.

5.2.2.1. Task difficulty. Table 5 clearly shows that the Written Review group was able to answer the questions faster than the Claim Network group. This was to be expected for several reasons. First the review has "added value" over the claim network: it was written by the researcher based on her understanding of the topics built up by constructing the claim network. Secondly, members of the Written Review group were far more familiar with the medium they were working with (essentially a reading comprehension test) than members of the Claim Network group. It was observed that all the members of the Written Review group used the printed version as their main resource and the version on the computer as an occasional look up. It would be unreasonable to expect similar ease of use with unfamiliar tools in comparison to a skill which the Written Review group had been practising for many years. Furthermore, the review was quite short, only about 2300 words. This was necessary to allow the Written Review group to read it and complete the questionnaire under experimental conditions. However, it is possible that if the review had been longer (e.g. 23,000 rather than 2300 words) the search and exploration services available to the Claim Network group would have given them an advantage.

The variability in the times taken by the Claim Network group was far greater than for the Written Review group. This seemed to be largely due to their personal style of question answering, in particular, the slowest participant had a very analytical approach to both the questions and the data in the claims.

We also observed that the Claim Network group generally gave far more "thinking aloud" contributions, which was an additional distractor and tended to increase actual time to answer questions. The reluctance of the Written Review participants to think aloud may stem from the strong habit of reading silently. Breaking that silence to comment on questions is a barrier.



Correctness

Network A	0	2	1	1.5	0	0	2	1.5	0	0	1.5
Network B	2	2	1.5	1.5	0.5	0.5	2	1.5	0	1	1
Network C	2	2	1.5	0.5	1	0	2	2	1	2	1.5
Review A	2	2	0.5	1	2	0	2	1	0	0.5	0.5
Review B	2	2	1	1	2	0.5	2	0.5	0	1	2
Review C	2	2	2	1	2	0.5	2	0.5	1	2	2

Fig. 16. Relative time to answer questions and correctness score per question.

Finally, the difference in correctness between the two groups' answers was small; the review group scored 1.4 more on average, a difference of about 6%. This difference is not particularly large, suggesting that the claim network can be understood by people other than the creator.

5.2.2.2. Per-question analysis. Fig. 16 shows in the bar chart the proportion of question answering time spent for each question and in the table the score of each participant on each question. For the review participants consideration was given to the fairest way to deal with reading time. Alternative approaches we considered were: to ignore it and start timing at the first point at which the participant started to address a question, to divide reading time equally between the questions (which seemed intuitively wrong), or to portion out the reading time across the questions in proportion to the amount of time spent on questions (which results in no change compared to ignoring reading time). The decision was complicated by the fact that only one of the group began the exercise by reading the review all the way through while other two started to answer some questions before they had finished reading the whole document. Consequently, we decided to opt for the first and simplest method which has the additional advantage of being most directly comparable with the times for the Claim Network group who did not have a reading period. These were calculated to remove the effect of personal style; for example, within the Claim Network group actual times were very variable. Relative time per question provides an indicator as to whether members of one group found certain questions relatively harder to answer than the other group.

For most of the questions there is no indication from performance times that one group of participants found any question noticeably harder than the other group. For question 4 the Written Review group found it easier to answer the question than the Claim Network group, whereas for questions 7 and 9 the latter completed the task in a relatively shorter time. In terms of correctness, only question 5 shows a difference between the groups with the review group all getting perfect scores and the Network group all scoring 1 or below. We will look at these three questions in detail.

The Claim Network group found question 4 (‘‘What is scientometrics?’’) relatively harder to answer than the Written Review group. The latter had little trouble finding a definition in the first sentence of the section of the review headed ‘‘scientometrics’’ and all copied this into the answer sheet as ‘‘*Scientometrics is the study of scientific research literature using citation data*’’. The Claim Network group took a much more exploratory approach. A similar definition had been embedded in the *notes* field of the node labelled ‘‘scientometrics’’. However, none of the Claim Network group thought to look for it. They knew that *notes* existed but most of the nodes did not have them and there was no visual flag to indicate that this node did (this is a user interface design flaw highlighted by the study). Instead they all looked at the nodes immediately adjacent to "scientometrics" and constructed their answers from those. These produced some acceptable definitions, e.g.

> From what I understand, scientometrics is a kind of meta-science, or a discipline, or a research area, that measures and represents 'discourse' phenomena within scientific research: for instance, providing a picture of how discourse develops in a field, through mapping literature (relations between researchers' work, perspectives, concepts, discourse acts like publications, etc.). To use ClaiMaker phrasing, it aims at providing evolutionary models of science, technology and scholarship.

This definition certainly reflects the background of the person who wrote it (a student of semiotics and discourse) but she has clearly formed a personal view of what scientometrics is. To summarize, while the process was more time consuming, it could be argued that the members of the Claim Network group were forced to engage more with the material and understand it, not having the option to simply paste the opening sentence conveniently found at the start of the Written Review.

By contrast it was the Written Review group who had trouble with question 7 ("Name three properties you would expect to see in social networks?"). The Claim Network group were helped by the text of three nodes. They all searched for nodes that contained the words "social networks" and got these: "*Social networks are assortative*", "*Social networks have a high degree of clustering*" and "*Social networks are divided into groups and communities*", as shown in Fig. 17.

Two out of three of the Claim Network group turned these directly into three properties with which to answer the question. The third decided that clusters were similar to groups and searched in the neighbourhood of these claims to extract the property: "*in social networks connections don't develop randomly*".

The Written Review group had greater difficulty in picking similar properties from the text. Each had to spend several minutes reading through the section of the text headed "Social networks" and one of them reported difficulties with the question. At the end each Written Review participant produced a rather different list of properties and only one of them gave a set that was similar to those reported by the Claim Network group. The Written Review group had to interpret several paragraphs of text in order to pick out properties whereas the Claim Network group had an easy way to answer the question. These differences come from the way pieces of information were presented in the two artefacts.

It is of course possible to design either artefact to highlight particular information, an issue to which we return in the discussion. However, the relative ease with which the Claim Network group tackled Question 9 ("Give the titles of two papers which report on combining information from Web pages with link analysis algorithms") stems from the generic affordances of the online system. Each Concept is related to a paper whose bibliographic details are stored within the system. An icon is presented with each concept that allows the user to open up a "details" box with the reference in it. Having identified relevant concepts the participants simply had to

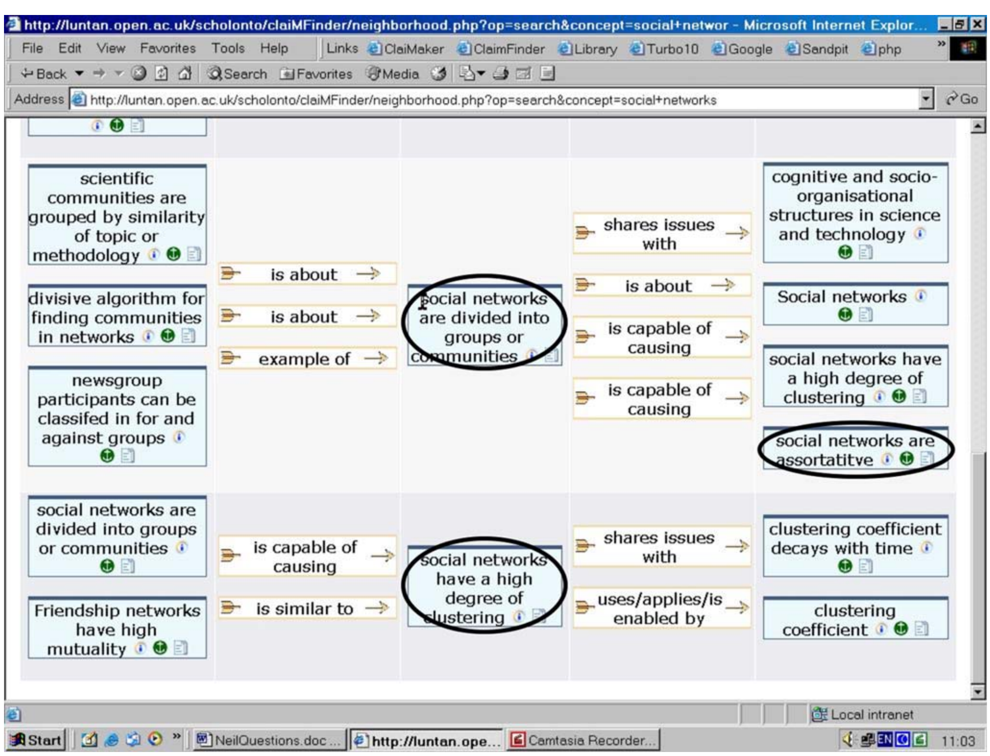


Fig. 17. Typical output from a search for "social networks" used to answer question 7.

check the details boxes and extract two different titles. The review participants had to deal with the familiar difficulties of matching a reference marker to the reference itself in a text document, although they mitigated this problem by doing text searches on authors' names in their online versions.

Finally, we will look at question 5 ("What does van Raan consider to be the sub-tasks of scientometrics?"), the only question for which there was a noticeable difference in scores between the two groups. In this case this difference can be explained by features in both artefacts. The written review had a bullet pointed list summarising van Raan's analysis. ClaimFinder, on the other hand, only had facilities for searching on content words not on authors' names. This had not posed a problem for the testers because they were both fluent users of ClaiMaker and simply switched over to the "expert" interface to answer this question. This difference reveals a design flaw in the novice interface that can be corrected by a simple change in how the ClaimFinder index is built.

5.2.2.3. Tackling interpretive questions. The first nine questions were fact-finding tasks. The last two questions required the participants to consider the material they had been presented with as a whole. Question 10 ("What unifying notion is common to scientometrics, social networks studies and link ranking algorithms?") required a synthesis of ideas encountered in answering the previous nine questions. The two groups showed quite different approaches to tackling this question. Two of the Claim Network group assumed there was some mechanism for tracing a path between concepts in the tool. In fact this facility did not exist and they ended up doing extensive searches of the neighbourhood of each concept trying to "*manually*" identify a Concept that was linked to all of them. One participant gave up when he could not find such a concept. The other two found the Concept "*Cognitive and socio-organisational structures in science and technology*" which is joined directly to both the "scientometrics" and "social networks" Concepts (Fig. 18). This helped each of them to start formulating an answer.

The Written Review participants knew there was no "magic button" they could press and that they would have to generate an answer from their own interpretation of the review. Their answers all drew on the idea of clustering and communities, ideas which had been mentioned in several sections of the review. Thus, on a high level, both groups produced answers about the identification of patterns, but in slightly different ways.

Question 11 is perhaps the most fundamental of the questions we asked and the most open-ended. Would the participants be able to identify open research questions using the information in the artefacts? As research students they had all been engaged in identifying research questions in their own domains but none of them specialized in the specific topics addressed in this study. Nor would one normally expect students to start formulating research questions after only being exposed to a domain for an hour or so. Therefore we did not expect particularly well-developed replies. We also expected that although the

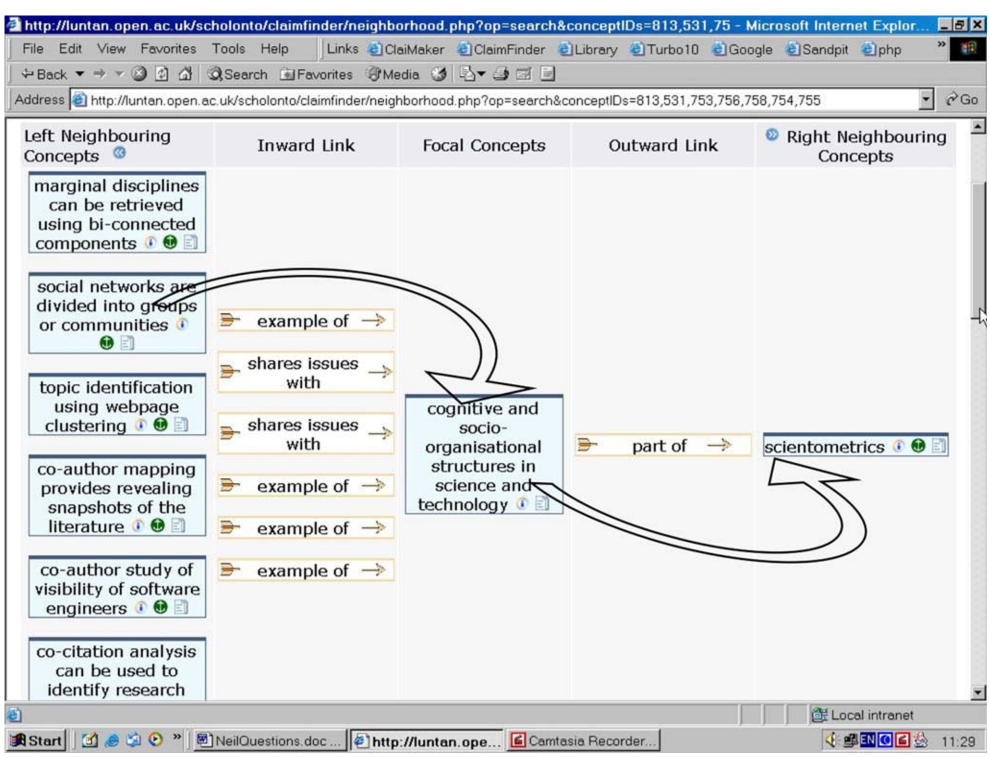


Fig. 18. Identifying a bridging concept in the ClaiMaker *Neighbourhood* visualization, in order to find a connection between two research fields (Question 10).

Table 6
Research questions identified by participants in the study in response to question 11

Participant	Research question proposed in response to question 11
A (Claim Network)	I'd have a look at the problems or shortcomings identified in one of these areas. For instance: I could work on clustering decays and try to reduce them. They are an important part of social networks and their reduction could increase the deployment of social network. Why? Because someone has said that they were an issue for social network. So why not tackle it?
B (Claim Network)	If my criteria was based on the amount of information available, I would have to choose the social networks aspect. This seemed to have been the best-covered topic in the database. In terms of personal interest and background knowledge I would have to choose the scientometrics aspect. The aspect I would least likely undertake would be the link ranking algorithms since this has a lot of terms that are not very familiar and not well defined in the database.
C (Claim Network)	If I wanted to study social networks on the web, I would try to look into the way people use links and express patterns of link use (paths). This would interest me because I think it would help me to identify people's thinking, and the way they interpret what they find, through the series of connections that they follow.
D (Written Review)	Social networks probably because it would inform the others.
E (Written Review)	I would be interested in research on social networks that evolve over time and that can be used for providing estimations of the importance of documents. Social networks and small-networks in particular appear to have attracted the interest of many researchers from various disciplines.
F (Written Review)	All of these methods rely on exploiting explicit links between papers. What appears to be missing is the reason for the reference. Once the network has been built and displayed graphically, it may be possible to use deeper NLP techniques to identify types of reference. A reference may be given for many reasons such as identifying the originator of some notion, theory or claim. A reference may be given because the particular work fills a gap revealed by another or it contradicts the other claim. It may be possible to deduce the nature of the citation on the fly when a reader is interested. Alternatively, this could be done for each citation which would be computationally expensive, but demonstrate different structures such as the genealogy of ideas or controversies, etc.

answers would be found partly in the material we presented, they might also relate to the students' own characters, interests and experience, since researchers' worldviews affect which questions they choose to ask and even find meaningful (Reich, 1994).

As a consequence of these factors we did not expect the participants to give similar answers to Question 11. Our aim was first and foremost to see if they could give any answers and secondarily to make a judgment of how useful their answers were. The answers the students gave are presented in Table 6.

All but one of the participants (D) produced a fairly complex answer. Participant D's answer, "*Social networks probably because it would inform the others*", is certainly not a foolish one. The student has realized that the papers researchers choose to cite and the pages they choose to make Web links to are partly social behaviours and that this could be an interesting thing to look at.

Some of the other answers demonstrate misconceptions on the part of the participants. For example, in A's question, the statement "I could work on clustering decays" comes from a search to look for problems that had been identified in the network as a stimulus for forming a research question. This brought up, among others the Concept "*clustering coefficient decays with time*", the word "decays" has negative connotations generally, but in this case it refers only to the decrease in a numerical measure of clustering observed in social networks as they grow. While it would be possible to study its causes, it is unlikely that the decay could be influenced significantly. Errors such as this and our observation of the candidates' behaviour in answering questions leads us to believe that if claim networks are to work well great clarity of language will be required. For the model we tested, the network contained very little text compared to the written review (although it is possible within the ScholOnto framework to attach detailed descriptions to nodes). This means that the understanding of the participants rested heavily on a few words. If there was an ambiguity, or if they were unfamiliar with the technical vocabulary in a field they could easily form a false opinion. This was less common for the Review group. Ambiguity itself is an issue relating to the content of the network model rather than any affordances of the tools, but there are interesting future challenges in building tools which support users in expressing themselves as clearly as possible.

Participant C makes assumptions about social networks research that are not true, probably reflecting her background in hypertext which forms a view of what properties a link has.

The influence of the students' backgrounds was a strong factor in the kinds of questions they chose. For example participant E is interested in adaptive algorithms, while F is studying natural language processing.

Four of the participants (A, C, D and F) provide answers which go beyond the information in the artefacts. They have clearly realized that to formulate a research question they will have to go beyond what has already been

Table 7
Number of search actions by type

Action type	Actions per person			Total actions	Expert user
	A	B	C		
ClaimFinder—Find	12	18	17	47	2
ClaimFinder—Discovery	3	9	5	17	5
ClaimFinder—Advanced	2	0	8	10	0
Total ClaimFinder searches	*17*	*27*	*30*	*74*	*7*
ClaiMaker searches	6	0	0	6	20
Icon—Anchor	22	34	52	108	35
Icon—Bibliographic	2	4	37	43	11
Icon—Concept	2	25	5	32	10
Total Icon led searches	*26*	*63*	*94*	*183*	*56*
Other search actions	0	0	0	0	9

done. Since they are split equally between the two groups it can be argued that the claim network was at least as good at communicating the information needed to start formulating research questions as the review. However the small number of participants restricts the generality of conclusions which can be drawn without further investigation.

5.2.2.4. Tool usage patterns in the Claim Network group. Finally we used the recordings of the Claim Network group's sessions to assess how the functions of the search tools were used. The numbers of search actions performed by the three participants were noted and divided into those using three different ClaimFinder search interfaces (*Find*, *Discovery* and *Advanced*), those using the ClaiMaker system and those using three of the icon links illustrated in Fig. 6. Search action data is presented in Table 7 and Fig. 19. In addition to the actions of the three participants, we have included in Table 7 a summary of the search actions of an expert user (the second author) who tested the questionnaire prior to the main experiment.

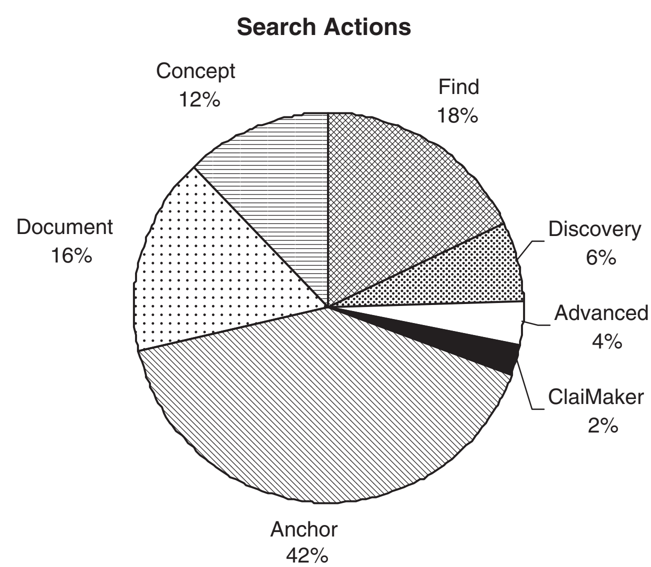


Fig. 19. Breakdown of search actions by type (totals for Claim Network Group).

The most heavily used features were the *Find* search in ClaimFinder (18%), a simple search for keywords in concepts and the *Anchor* icon (42%), which selects a concept to be the focus of a *Neighbourhood* search, as described in Section 4.2.3. This pattern of use reflects the dominant searching strategy, which was to perform a keyword-based search to locate the topic required and then to explore the local region of the network.

All three participants used some of the *Discovery* and *Advanced* search features in ClaimFinder. This was sometimes because the dominant *Find/Anchor* pattern had failed, but another motivation seemed to be simple exploration; as they grew more used to the results of the simple *Find* they explored new techniques.

Only one participant (A) used any of the ClaiMaker features even though all the participants were shown that shortcuts to both were given on the toolbar. Participant A was a research student who had been involved in developing input tools for ScholOnto models and had previous experience of using ClaiMaker. He used it briefly at the beginning of his session before concentrating on the more attractive ClaimFinder interface.

Participants B and C showed a bias towards using the *Concept* and *Document* icons, respectively. Both used them mainly for checking bibliographic data, which is duplicated in the two places. This may merely reflect habits established by early success with one method or the other. It perhaps indicates that the two icons could be merged to reduce clutter in the display.

When we compare the actions taken by the three students with those of the expert user (Fig. 20) we saw a similar foraging behaviour with use of an initial service followed by repeated use of the *anchor* icon for exploration. As would be expected, the expert has a much wider repertoire of actions and the initial service was not usually the *Find* service in ClaimFinder.

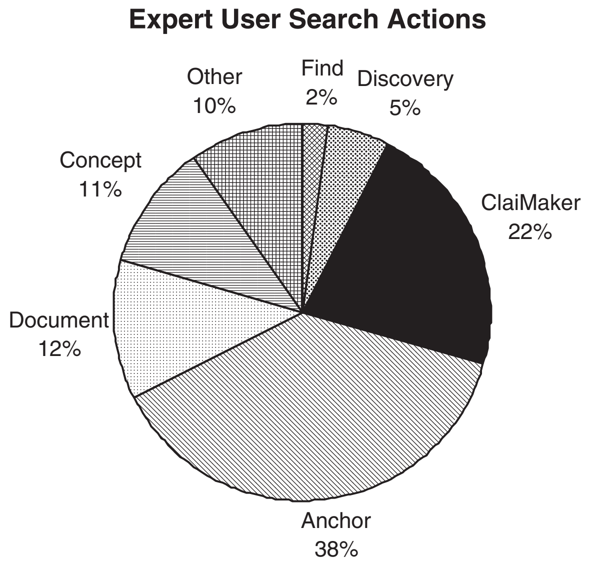


Fig. 20. Breakdown of expert user search actions.

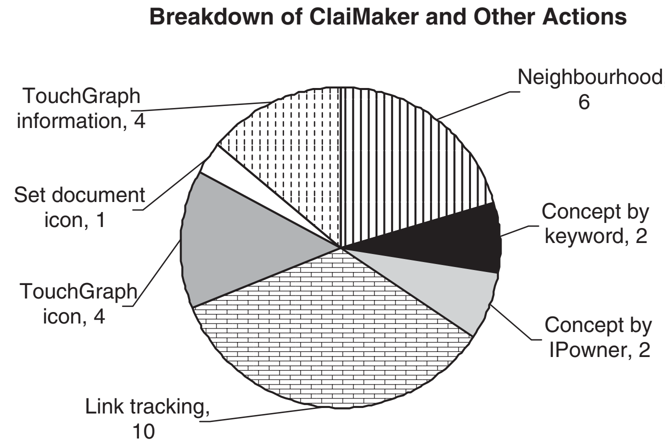


Fig. 21. Breakdown of ClaiMaker and other search actions for the expert user.

He used the "unfriendly" ClaiMaker a lot more, 22% of actions compared to 2% for the students. This may be partly because the expert had more than 2 years experience of using ClaiMaker so tended to turn to it before the less familiar (to him) ClaimFinder interface. However there is evidence that he was exploiting services and controlling parameters via the ClaimMaker interface that were not available through ClaimFinder. Fig. 21 gives a breakdown of the ClaiMaker actions of the expert user and the actions included in the "other" category.

Some of the expert's ClaiMaker actions mirror the ClaimFinder usage of the Claim Network group. He used *Neighbourhood* six times, which gave him the same output as the *Find* service in ClaimFinder, but more control over the parameters he could search for, e.g., he could search for words in the title of an article as well as keywords in Concepts. His searches for concepts by keyword and IP Owner mirrored the kind of services that can be found through the ClaimFinder *Advanced* menu.

His use of *link tracking* (see Fig. 10 and screen movie clip[8]), however, shows him using a service which was not put into ClaimFinder because it was considered too complex for a novice user. It gives the user a lot of options, including keywords in left- or right-hand concepts, specific link types, groups of link types drawn from the taxonomy and the depth of search. Having done a *link-tracking* search the expert user would then often explore the TouchGraph visualization of the results, by clicking the TouchGraph icon, a form of presentation which the students did not use heavily.

5.2.2.5. Summary of comparative evaluation study. To summarize, while there was a clear advantage in terms of actual time taken for the Written Review group, it could be argued that the review had added value over the Claim Network. Furthermore the Written Review group had a major advantage in terms of experience with the artefacts. Both groups gave appropriate answers to the questions suggesting that the claim network was interpretable by users other than its creator. In terms of tool use, the three Claim Network participants concentrated mainly on using the simpler functions. However the ClaimFinder tool seemed to invite them to use more complex functions as their confidence increased. The expert user's usage patterns suggest that, with increased 'literacy' with these new tools, users develop more complex search strategies.

6. Conclusions

New technologies, such as digital libraries, have increased the availability of documents dramatically. For researchers this has generated a need for better tools to make sense of the many papers they can now access. In the Scholarly Ontologies project, we proposed a computational approach to support such scholarly sensemaking. Our argument is that classical truth maintenance models would not be fit for this purpose. Instead the scheme adopted must enable evidence to be presented simultaneously in favour of claims and complemented by counter-claims. Thus we propose an ontology of rhetorical relations for principled agreement and disagreement which can support multiple interpretations. This uses a claim network representation to model a document's key contributions and relationships to the literature. The network approach has focussed our knowledge modelling effort on capturing relationships between objects, rather than simply indexing instances of objects. To this end the Scholarly Ontologies project has investigated a new kind of digital library server in which it would be possible to go beyond searching metadata and to ask questions more pertinent to research.

In this paper, we have presented three related prototype systems which have been developed during the project to support users in the creation and exploration of claim

[8]A movie clip of the expert user performing a link-tracking search as part of this study is available at http://claimaker.open.ac.uk/.

networks, namely ClaiMaker, ClaiMapper and ClaimFinder. We also presented a two-phase evaluation study in which a review was made of a multi-disciplinary domain during which a claim network was built. Two groups of students, one group using a written literature review based on this modelling and the other using the claim network itself, answered questions about the domain. This case study allowed us to investigate the utility of the tools and the approach. The results of the case study indicate that the claim network approach presented here can support services which address questions like "where did this idea come from?" and "what evidence supports this idea?" Furthermore, it was observed that as the review progressed the degree of order in the models produced in the ClaiMapper interface increased, suggesting that it was able to support and reflect back to the analyst, the process of conceptual refinement, an important part of sensemaking activity.

The case study has demonstrated that claim networks helped the participants to form their own opinions. For instance, the Claim Network participants were more inclined than the Written Review group to construct personal answers rather than to extract answers from the artefact as if they were "truth". Possibly this is because the relative sparseness of the network representation forces the user to mentally fill in some of the gaps in the information they are presented with, or it may simply be that dealing with an unfamiliar representation forces the user to think harder, with a richer articulation of ideas as a positive side-effect. While it is beyond the scope of this study to speculate further on the cognitive processes of the participants, it is worth noting that working with the claim networks brought out some different thinking skills in line with positive results reported for the use of argument visualization in teaching reasoning skills (Carr, 2003) and for mind mapping in education (Novak and Gowin, 1984).

In addition, we observed a common search strategy in which the students using the claim network performed a simple search to locate a node with the right keywords in it and then explored the region of claims around it, often using the anchor icon to initiate neighbourhood searches. This behaviour is typical of the information foraging behaviour described by Pirolli and Card (1999). The users are locating what is called an "information patch" and then grazing on the information in the patch until either they can answer the question to their satisfaction or they have exhausted the information there and need to move to a new patch. Although the data from this case study is limited to only three users, all of them demonstrated foraging behaviour, which is common in information systems generally. It seems reasonable to conclude that it was an important way for them to interact with these models. The next round of tool development should therefore focus on developing the functionality of browsing tools and the clarity of their outputs to help users forage as effectively as possible. ACT-IF, the formal process model presented by Pirolli and Card in their 1999 paper to describe information foraging, presents a candidate cognitive modelling approach to evaluate new tools or interfaces aimed at supporting browsing.

The sparseness of the representation had disadvantages; when the representation was ambiguous it could cause misunderstanding. It seems that if claim networks are to be used collaboratively the users will have to learn to express themselves precisely, which is to say in the terms that their community will understand. Ambiguities may also emerge as triggers for debate when communities start working together on building networks.

Finally, while it was demonstrated that the claim networks can convey information about a subject domain, we do not claim it is a substitute for text. Text allows the author more flexibility in constructing the narrative and more influence on the reader because the author has control of the order in which information is presented. Our observations suggest that our participants could handle information in written format faster than using the claim network. That said, the prose review and claim network were both relatively small and the Claim Network group had much less experience of the medium than the Written Review group. It is possible that if the written review had been longer and the users of the claim network had more experience, it would have been possible to better show the benefits of the search services.

In summary, the ScholOnto research project has been envisioning how scholarly knowledge may be published and contested in the future. A variety of prototype tools have been developed in our pursuit of an environment which would enable analysts to express their perspective of the ideas in a literature, which could then be published and interrogated as a personal, or shared, model. The evaluation reported has taken the first step by demonstrating that given the right tool, literature analysis can be assisted by the construction of claim networks and that, within the limits of the study reported, these networks could effectively convey information to users other than their creator. Although the tools we produced support group-working at a technical level, we have not yet studied their synchronous or asynchronous use in collaborative environments. We conclude that the Scholarly Ontologies approach to sensemaking is worthy of further investigation, to improve and further evaluate the tools.

Acknowledgements

We gratefully acknowledge the support of the UK Engineering and Physical Sciences Research Council's Distributed Information Management Programme (GR/N35885/01), 2001–2004.

References

Berners-Lee, T., Hendler, J., Lassila, O., 2001. The semantic web. Scientific American, 34–43.

Buckingham Shum, S., 2003. The roots of computer supported argument visualization. In: Kirschner, P.A., Buckingham Shum, S., Carr, C. (Eds.), Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making. Springer, London, pp. 3–24.

Buckingham Shum, S. (Ed.), 2005. Scholarly Hypermedia, New Review of Hypermedia and Multimedia (Guest Editorial, Special Issue on Scholarly Hypermedia) 11 (1), 1–6.

Buckingham Shum, S., Hammond, N., 1994. Argumentation-based design rationale: what use at what cost? International Journal of Human-Computer Studies 40 (4), 603–652.

Buckingham Shum, S., Motta, E., Domingue, J., 2000. ScholOnto: an ontology-based digital library server for research documents and discourses. International Journal on Digital Libraries 3 (3), 237–248.

Buckingham Shum, S., Uren, V., Li, G., Domingue, J., Motta, E., Mancini, C., 2002. Designing representational coherence into an infrastructure for collective sensemaking. In: Proceedings of the Second International Workshop on Infrastructures for Distributed Collective Practices, San Diego.

Buckingham Shum, S., Uren, V., Li, G., Domingue, J., Motta, E., 2003. Visualising internetworked argumentation. In: Kirschner, P.A., Buckingham Shum, S., Carr, C. (Eds.), Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making. Springer, London, pp. 185–204.

Bush, V., 1945. As we may think. The Atlantic Monthly 176, 101–108.

Buzan, T., 1989. Use Your Head. BBC Books, London.

Carr, C.S., 2003. Using computer supported argument visualization to teach legal argumentation. In: Kirschner, P.A., Buckingham Shum, S.J., Carr, C.S. (Eds.), Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making. Springer, London, pp. 75–96.

Cronin, B., 2001. Bibliometrics and beyond: some thoughts on web-based citation analysis. Journal of Information Science 27 (1), 1–7.

Duncan, E.B., Anderson, F.D., McAleese, R., 1981. Qualified citation indexing: its relevance to educational technology. In: Information Retrieval in Educational Technology: Conference Proceedings of the First Instruments of Cognition 27 Symposium on Information Retrieval in Educational Technology, ETIC'81, University of Aberdeen, Aberdeen, Scotland.

Engelbart, D.C., 1962. Augmenting Human Intellect: A Conceptual Framework. Stanford Research Institute.

Garzone, M., Mercer, R.E., 2000. Towards an automated citation classifier. In: Advances in Artificial Intelligence: Proceedings of the 13th Biennial Conference of the Canadian Society for Computational Studies of Intelligence, AI 2000, Montréal, Quebec, Canada, May 2000. Springer, Berlin.

Gil, Y., Ratnakar, V., 2002. TRELLIS: an interactive tool for capturing information analysis and decision making. EKAW 2002, LNAI 2473.

Green, T.R.G., 1990. Cognitive dimensions of notations. People and Computers V: Proceedings of the British Computer Society HCI'89 Conference. Cambridge University Press, Cambridge.

Gruber, T.R., 1993. Towards principles for the design of ontologies used for knowledge sharing. In: Guarino, N., Poli, R. (Eds.), Formal Ontology in Conceptual Analysis and Knowledge Representation. Kluwer Academic Publishers, Dordrecht.

Handschuh, S., Staab, S., 2002. Authoring and annotation of web pages in CREAM. In: Proceedings of the 11th International World Wide Web Conference (WWW2002), Honolulu, Hawaii.

Kahan, J., Koivunen, M.-J., Prud'Hommeaux, E., Swick, R., 2001. Annotea: an open RDF infra-structure for shared web annotations. In: Proceedings of the 10th International Conference World Wide Web Conference (WWW10), Hong Kong.

Kamp, H., 1981. A theory of truth and semantic representation. In: Groendijk, Janssen, Stokhof (Eds.), Formal Methods in the Study of Language. Mathematisch Centrum, Amsterdam.

Kim, H.J., 2000. Motivations for hyperlinking in scholarly electronic articles: a qualitative study. Journal of the American Society for Information Science 51 (10), 887–899.

Kleinberg, J.M., 1999. Authoritative sources in a hyperlinked environment. Journal of the ACM 46 (5), 604–632.

Knott, A., Mellish, C., 1996. A data-driven method for classifying connective phrases.

Knott, A., Sanders, T., 1998. The classification of coherence relations and their linguistic markers: an exploration of two languages. Journal of Pragmatics 30, 135–175.

Lassila, O., 2001. Enabling semantic web programming by integrating RDF and common lisp. In: Proceedings of the SWWS, Semantic Web Working Symposium, Stanford.

Lipetz, B.-A., 1965. Improvement of the selectivity of citation indexes to scientific literature through inclusion of citation relationship indicators. American Documentation 16 (2), 81–90.

Mancini, C., 2005. Towards Cinematic Hypertext: A Theoretical and Empirical Investigation. IOS Press, Amsterdam.

Mancini, C., Buckingham Shum, S., 2004. Towards 'Cinematic' Hypertext. In: Proceedings of the 15th ACM Conference on Hypertext & Hypermedia, Santa Cruz, CA, USA.

Mercer, R.E., Di Marco, C., 2004. A design methodology for a biomedical literature indexing tool using the rhetoric of science. Linking Biological Literature, Ontologies and Databases, HLT-NAACL 2004 Workshop: Biolink 2004, Association for Computational Linguistics.

Murugesan, P., Moravcsik, M.J., 1978. Variation of the nature of citation measures with journals and scientific specialties. Journal of the American Society for Information Science 29 (3), 141–147.

Nanba, H., Okumura, M., 1999. Towards multi-paper summarization using reference information. In: Proceedings of the 16th International Joint Conferences on Artificial Intelligence (IJCAI '99), Stockholm, Sweden.

Novak, J.D., Gowin, D.B., 1984. Learning How to Learn. Cambridge University Press, Cambridge.

Page, L., Brin, S., Motwani, R., Winograd, T., 1998. The PageRank Citation Ranking: Bringing Order to the Web. Stanford University.

Pirolli, P., Card, S.K., 1999. Information foraging. Psychological Review 106 (4), 643–675.

Pool, I. de Sola, Kochen, M., 1978/79. Contacts and influence. Social Networks 1, 5–51.

van Raan, A.F.J., 1997. Scientometrics: state-of-the-art. Scientometrics 38 (1), 205–218.

Reich, Y., 1994. Layered models of research methodologies. Artificial Intelligence for Engineering Design, Analysis and Manufacturing (AI EDAM) 8 (4), 263–274.

Selvin, A.M., Buckingham Shum, S., 2005. Hypermedia as a productivity tool for doctoral research. New Review of Hypermedia and Multimedia (Special Issue on Scholarly Hypermedia) 11 (1), 91–101.

Sereno, B., Buckingham Shum, S.J., Motta, E., 2005. ClaimSpotter: an environment to support sensemaking with knowledge triples. In: Proceedings of the IUI2005: ACM Conference on Intelligent User Interfaces, San Diego. ACM Press, New York.

Shipman, F.M., Marshall, C.C., 1999. Formality considered harmful: experiences, emerging themes, and directions on the use of formal representations in interactive systems. Computer Supported Collaborative Work 8 (4), 333–352.

Teufel, S., Moens, M., 2000. What's yours and what's mine: determining intellectual attribution in scientific text. In: Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora.

Teufel, S., Moens, M., 2002. Summarizing scientific articles: experiments with relevance and rhetorical status. Computational Linguistics 28 (4).

Toulmin, S., 1958. The Uses of Argument. Cambridge University Press, Cambridge.

Trigg, R., 1983. A Network-based Approach to Text Handling for the Online Scientific Community. Department of Computer Science, University of Maryland.

Uren, V.S., Buckingham Shum, S., Li, G., Domingue, J., Motta, E., 2003. Scholarly publishing and argument in hyperspace. In: Proceedings of the 12th International World Wide Web Conference, WWW2003, Budapest, Hungary. ACM Press, New York.

Uren, V.S., Buckingham Shum, S., Mancini, C., Li, G., 2004. Modelling naturalistic argumentation in research literatures. In: Proceedings of the Fourth Workshop on Computational Models of Natural Argument, held in conjunction with ECAI 2004: European Conference on Artificial Intelligence, Valencia.

Weick, K.E., 1996. Sensemaking in Organizations. Newbury Park, CA, Sage.

Weinstock, M., 1971. Citation Indexes (part 1). Encyclopedia of Library and Information Science 5, 16–40.

White, H.D., McCain, K.W., 1989. Bibliometrics. Annual Review of Information Science and Technology (ARIST) 24, 119–186.