PrePrint: To appear: *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*. Paul A. Kirschner, Simon J. Buckingham Shum and Chad S. Carr (Eds), pp. 185-204, In Press. Springer-Verlag: London [www.VisualizingArgumentation.info] (Publication due in December 2002)

Visualizing Internetworked Argumentation

Simon Buckingham Shum, Victoria Uren Gangmin Li, John Domingue, Enrico Motta

Knowledge Media Institute, Open University, UK

9.1 Scholarly Publishing and Argumentation: Beyond Prose

In this chapter, we outline a project which traces its source of inspiration back to the grand visions of Vannevar Bush (scholarly trails of linked concepts), Doug Engelbart (highly interactive intellectual tools, particularly for argumentation), and Ted Nelson (large scale internet publishing with recognised intellectual property). In essence, we are tackling the age-old question of how to organise distributed, collective knowledge. Specifically, we pose the following question as a foil:

> In 2010, will scholarly knowledge still be published solely in prose, or can we imagine a complementary infrastructure that is 'native' to the emerging semantic, collaborative web, enabling more effective dissemination and analysis of ideas?

We are neither trying to replace textual narrative as an expressive medium, nor its products such as books and peer reviewed publications. We seek instead to augment them by exploiting globally networked information in ways that – precisely because of its historical pedigree – the venerable prose publication cannot support. Conventional scholarly publications are the way they are through a co-evolution of notational form with print publishing technology, but are not designed in any way to take advantage of today's information infrastructure. Still at a relatively early stage, our project is bringing to bear on this challenge a *networked representational environment* (a digital library server based on an argumentation ontology (Buckingham Shum *et al.*, 2000), *semantic web services* (e.g. ontology-based reasoning, Li *et al.*, 2002), and recent work on *distributed collective practices* (why and when individuals in a community of practice are willing to subscribe to a shared repository, and role of formalism, Buckingham Shum *et al.*, 2002). All of these must be interacted with via a variety of *user interfaces*, of which a key component will be renderings of the network of *argumentative claims*—the focus of this chapter.

We start with some background needed to understand the rationale for this work, and refer the reader to other sources for more detailed treatments of technical and social issues. We then focus on challenges associated with designing visual interfaces onto a shared repository, in order to construct what we call 'claims' about research contributions. Finally, we conclude by outlining the agenda for future work.

9.2 What's the Problem?

Researchers are benefiting from more rapid access to research documents as resources such as new digital libraries and eprint archives go online almost by the week, but researchers (like almost all other professions) are also drowning in this ocean, with less time to track growing numbers of conferences, journals and reports. But beyond tracking new results, there is the whole dimension of analysing a literature. 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?* These are of course questions about the *meaning* of a research contribution. Such questions operate at a different level from that addressed by conventional metadata or ontological markup, which normally seek to iron out inconsistency, ambiguity and incompleteness (clearly undesirable for details such as bibliographic or other uncontentious details). In contrast, principled disagreement about significance, conflicting perspectives, and the resulting ambiguities and inconsistencies are precisely what define a field as research; they are the objects of explicit inquiry. It in this context that structured argumentation has a contribution to make. In sum, there remains a yawning gap in the researcher's digital toolkit: tools to track *ideas* and *results* in a field, and to express, analyse and contest their *significance*.

As well as characterising this problem, the Scholarly Ontologies (ScholOnto) project is developing a system to support scholarly interpretation and argumentation, investigating the practicality of publishing explicit conceptual structures (grounded in conventional documents) in a collective knowledge base. The *ClaiMaker* system enables researchers to make *claims*, that is, to describe and debate, in a network-centric way, their view of a document's key contributions and relationships to the literature. It thus provides an interpretational layer above raw resources (such as documents, datasets, and tools).

We hypothesise that this will be of value to a variety of end-users: filtered views for students onto major debates in their field (as pioneered on paper by Horn: Chapter 5); tools for information analysts/librarians to conduct literature analyses; alerting services for researchers working across inter-disciplinary boundaries in which it is impossible to track all relevant research; visual browsing of concept networks to locate relevant documents. We turn now to the argumentation scheme that underpins the making of claims in the system.

9.3 The Discourse Ontology

"Ontologies" are the term used in knowledge modelling and agent research, and increasingly within the semantic web community, to describe an abstract (implementation-independent) specification of concepts, attributes and relationships

(Gruber, 1995). Typical semantic web work develops an ontology to control interpretation or semantic annotation in a specific domain of inquiry (such as an ontology of problem-solving methods) or to model a particular aspect of the world (such as organisational functions), enabling machine-to-machine interoperability and interpretation. In contrast, we propose an ontology for scholarly discourse, primarily for *humans* to communicate through as a medium for publishing and discourse (although we envisage agents as protagonists and claim-makers at some point), with the express goal of supporting multiple (often contradictory) perspectives. In this sense it is as much an ontology for *principled disagreement*. Of course, it requires consensus in the sense that participants subscribe to the ontology as a reasonable language for "making and taking perspectives" (Boland and Tenkasi, 1995), but they need not agree at all on the actual issues under debate.

The requirements for the ontology that we aimed for are summarised in Table 9.1.

Table 9.1: Motivating requirements for the research discourse ontology.

Requirements for a scholarly discourse ontology
1. Mimic natural language expressions to reduce the cognitive gap. An underlying structure based on a noun/verb metaphor with the relations taking the role of verbs seemed appropriate. Making arguments in pseudo-natural language should make the scheme intuitive for contributors.
2. The scheme must permit the expression of dissent. The ScholOnto project is fundamentally about argumentation and, more broadly, scholarly discourse (not all of which is argumentative). The ontology is not there to impose a single domain model, but to support the contesting of *perspectives*.
3. Ownership of public content is critical. Contributors must take responsibility for the claims they make. ClaiMaker's content could be filtered via a formal peer review process, but in early versions we depend on the social control of peer pressure to motivate high quality claim-making. Ownership also has a key role in ClaiMaker as digital library server: claims would be "backed up" by a link to a published paper. There is an analogy here with Toulmin's (1958) warrants.
4. Social dimensions to being explicit. ClaiMaker invites researchers to consider making explicit what is normally implicit in the text of a paper (an issue discussed in Buckingham Shum *et al.*, 2000). Discourse relational types vary in strength, which has both computational and social dimensions. Consider a relation *refutes*. This is a forceful term and therefore can carry greater weight in computation than, for example, *takes issue with*. From a social perspective, 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. A concept has no category outside of use. 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 (or where an interpretation can be imposed), this cannot be sustained when we are talking about the *role* that a concept plays in multiple arguments in research: after all, an idea that is a *Problem* under debate in one paper may be an *Assumption* in another. The scheme must therefore allow the same concept to take on different types in different situations: meaning derives from context, where context is the forging of a connection between two ideas. It may even be impossible, or too much cognitive effort, to try and classify the concept (e.g. whether something is classed as a *Method*, *Theory*, *Language*, or

all three, may not be of real interest).

6. **The scheme should recognise disciplinary differences in argumentative style.** We are trying to identify a core set of argumentation relations that are useful in many disciplines. However, the precise language used for making a case will differ from one research community to another. We tackle this using the idea of *dialects*. Drawing on Cognitive Coherence Relations[1], 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* your colleagues could change the label to something they felt more comfortable with (e.g. *is inconsistent with; challenges; raises issues with*), but the notion of a negative relation that challenges a concept would remain unchanged. This method would let us configure ClaiMaker for different communities without altering the underlying engine.

Based on our intuitions as researchers, and drawing on related computational linguistics work on 'coherence relations' (Mancini and Buckingham Shum, 2001), plus earlier work on hypertextual argumentation (e.g. Newman and Marshall, 1991; Trigg and Weiser, 1983), a prototype discourse ontology was devised to satisfy this list of requirements. It had two basic object types: data and concept. The most important type of data object is a set of metadata describing a document in a digital library, these provided the backing, every claim being grounded in a published document (a quality-control policy decision – more open policies could be adopted). Concepts are stored as short pieces of free text succinctly summarising a 'contribution' (at whatever granularity the researcher wishes to express this), for instance: *<Data> Undergraduate chemistry exam performance is doubled after training on the ChemVR system*. This is now an object that others can connect to, whether positively or negatively. A claim is a triple (Figure 9.1) of two objects connected by a link.

Each link is drawn from a general class (e.g. *Problem-related; Taxonomic; Causal*), has the properties type, polarity and weight, and a dialect label in natural language. A concept may optionally be assigned a type (e.g. *Data, Language, Theory*), stored as part of the link connecting it. By storing the concept type in the link, rather than binding it intrinsically to the concept, the typing of concepts is made context dependent. Researchers may of course disagree on the concept's type, a common focus for discussion some fields (e.g. is this *Language* also a *Theory*? Is this based on *Opinion* or *Data*?).

1 Cognitive Coherence Relations (e.g. Knott and Mellish, 1996; Knott and Sanders, 1998) is a field in psycholinguistics which investigates the question of whether there is a core set of cognitive relationships that underpin written language. This field is summarised and related to the ScholOnto project by Mancini and Buckingham Shum (2001).

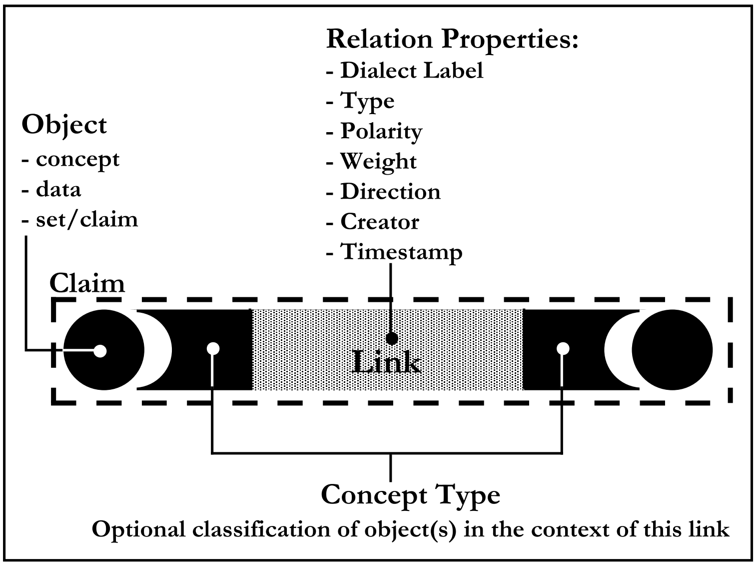

Figure 9.1: Structure of a *Claim* in the discourse ontology.

Elsewhere we have described the iteration from the first to the current version of the ontology (Buckingham Shum *et al.*, 2002), a process which itself was supported by the IBIS approach for capturing team deliberations (described by Conklin, Chapter 8 and Selvin Chapter 7). The current scheme is summarised in Table 8.2. Our goal is to provide a given research community with a dialect that will cover the most common claims that they make (there may well be exceptional kinds of contributions that fall outside the expressiveness of the vocabulary, but the generic *Other Link* is available for those situations). We are aware that the scheme could be much more expressive, rigorous and formal. However, as we discuss elsewhere (Buckingham Shum *et al.,* 2000), we are walking the tightrope between usability and formal rigour, and important lessons have already been learnt about over-formalizing interactive systems for untrained users (Shipman and Marshall, 1999).

To summarise, we propose that the kinds of connections shown in Table 8.2 are expressed at a level which most researchers would not only recognise, but indeed, would naturally use when summarising part of a literature[2]. Our internal testing shows that with a little practice, fluency in thinking in these terms is not hard to acquire, although of course, we will only know how generalisable this finding is as we study the system in wider use.

---

2 One strand of the ScholOnto Project is analysing the text of research publications, and shows some promise that we can indeed highlight phrases that correspond to claims in ClaiMaker. This would assist in the submission of new claims to ClaiMaker.

Table 9.2: The revised discourse ontology following a first iteration and use analysis.

Relation Class	Dialect label	Polarity/Weight
General	is about	+/1
	uses/applies/is enabled by	+/1
	improves on	+/2
	impairs	– /2
	other link	+/1
Problem Related	addresses	+/1
	solves	+/2
Supports/Challenges	proves	+/2
	refutes	–/2
	is evidence for	+/1
	is evidence against	–/1
	agrees with	+/1
	disagrees with	–/1
	is consistent with	+/1
	is inconsistent with	–/1
Causal	predicts	+/1
	envisages	+/1
	causes	+/2
	is capable of causing	+/1
	is prerequisite for	+/1
	prevents	–/2
	is unlikely to affect	–/1
Similarity	is identical to	+/2
	is similar to	+/1
	is different to	–/1
	is the opposite of	–/2
	shares issues with	+/1
	has nothing to do with	–/1
	is analogous to	+/1
	is not analogous to	–/1
Taxonomic	part of	+/1
	example of	+/1
	subclass of	+/1
	not part of	–/1
	not example of	–/1
	not subclass of	–/1

9.4 Making Claims Requires Mental Mapping

Given this underlying language, we now turn to the specific challenge of making claims structures visible in a coherent manner, whether at the point of creation, or when browsing/searching. We implemented the first *ClaiMaker* user interface as rapidly as possible in order to understand the authoring process, evaluate the ontology and populate the knowledge base. Now in its second main design iteration, this is a web forms/menu based design, is illustrated in Figure 9.2.

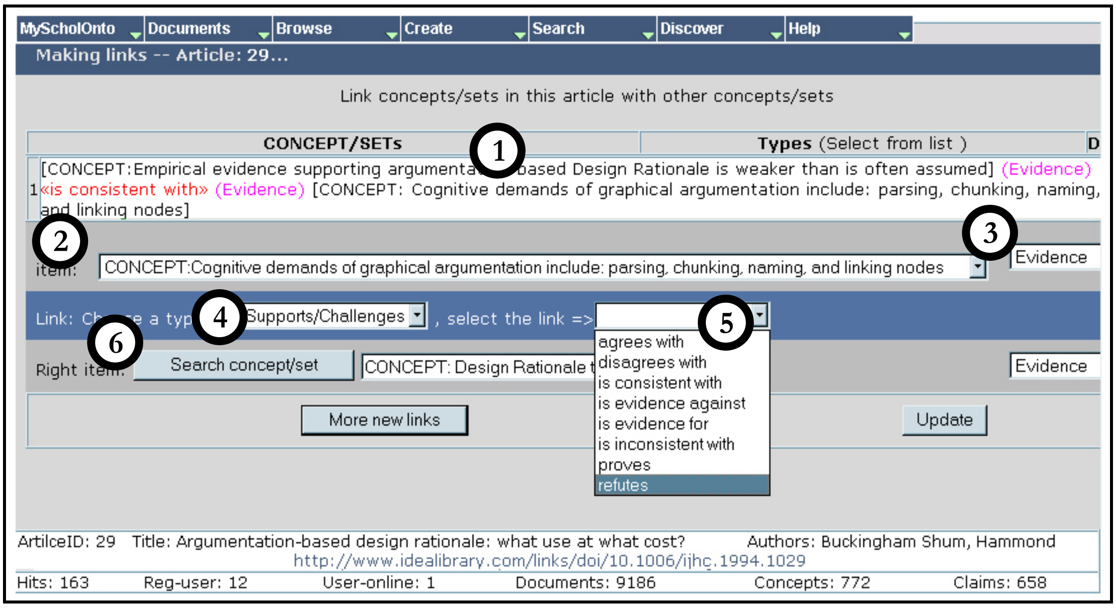

Figure 9.2: User interface to ClaiMaker, showing how a researcher can build a set of claims. Key: (1) A claim that has already been constructed, ready to submit; (2) the Concept to link from, which has (3) been assigned the type Evidence, and (4) linked via the Relational Class Supports/Challenges, (5) more specifically, refutes (selected from the dialect-specific menu). (6) The user then searched the knowledge base for a target Concept, Set or Claim to which they wish to make the connection.

Although this supports claim-construction at a technical level, and menu-based form-filling is a familiar activity to web users, we are also exploring a complementary user interface approach. Making claims is essentially literature modelling, a cognitive task that requires the mental construction of a network structure. In our experience, externalising this through conceptual maps (sketched or diagrammed in software) is the most intuitive way in which to manage the cognitive load, and leading typically, to iterative refinement of the model as concept names and types, link types, and granularity are revised (cf. earlier empirical studies into the cognitive demands of graphical argumentation, Buckingham Shum *et al.*, 1997). Figure 9.3 shows an example of the cognitive map that is typically produced when a researcher starts to think about how to describe a literature in terms of claims. It is also the kind of representation that one often constructs *prior to* using ClaiMaker, a signal that the tool is not supporting all phases of important cognitive work: to clarify one's thinking prior to adding claims in the forms/menus interface, one first sketches visually.

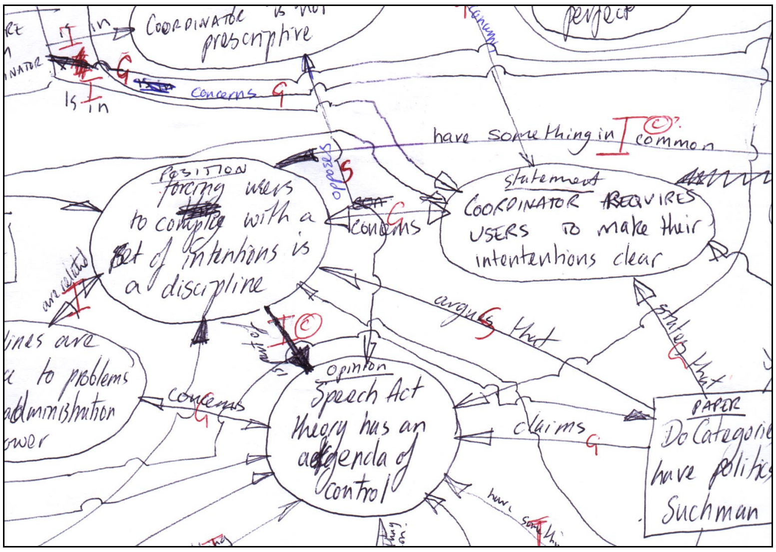

Figure 9.3: Sketching is a requirement for managing the cognitive task of modelling complex claim structures in a literature.

We are now developing a concept mapping user interface, screens from an early version of which are shown in the following section.

9.5 Visual Construction of Argumentative Claims

In the spirit of practising what we preach, let us take as a local example: the very book that you are now reading. What would it mean to represent the key contributions of each chapter, and the connections (both inter-concept, inter-chapter and to the roots and wider literature in the field) as an explicit claims network of concepts and associated argumentation? What kind of user interface could we provide to map out this structure, and what representational issues arise in the process?

Figure 9.4 shows a claims analysis of Chapter 7 by Selvin[3]. We have adopted a vertical layout convention, with a primary concept at the top ("primary", of course, by our reading of the chapter, and in order to make a specific point with our map; different readers might produce different maps). Under this we unfold the supporting argument that is presented.

3 The modelling has been done in an adaptation of the *Mifflin* tool for IBIS argumentation <www.compendiuminstitute.org/tools/mifflin.htm>

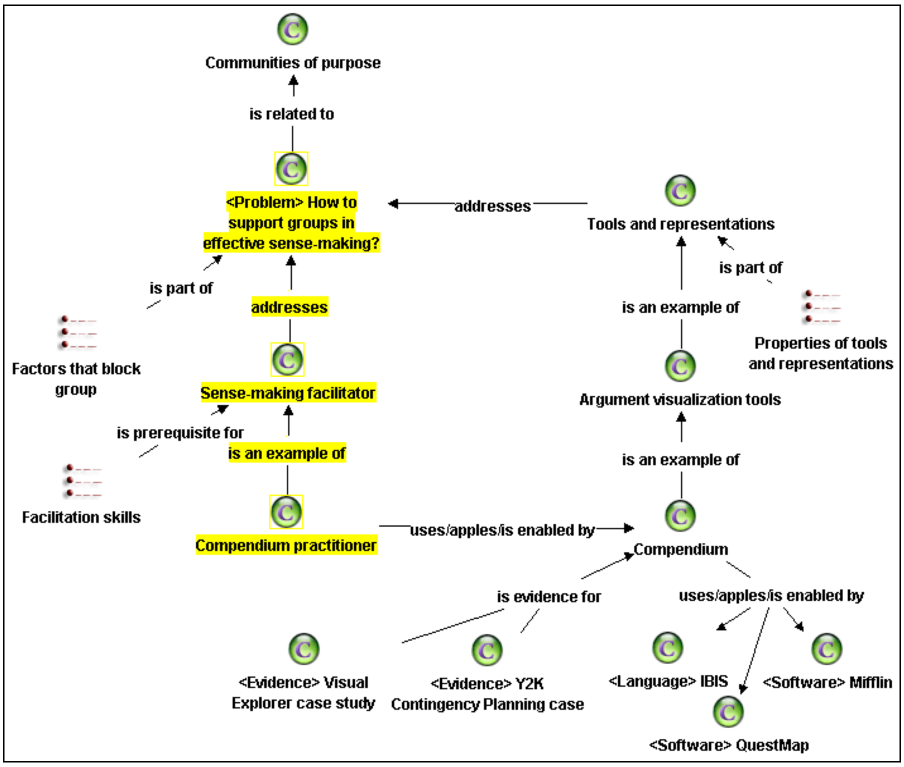

Figure 9.4: Visual claims analysis of part of Chapter 7 by Selvin.

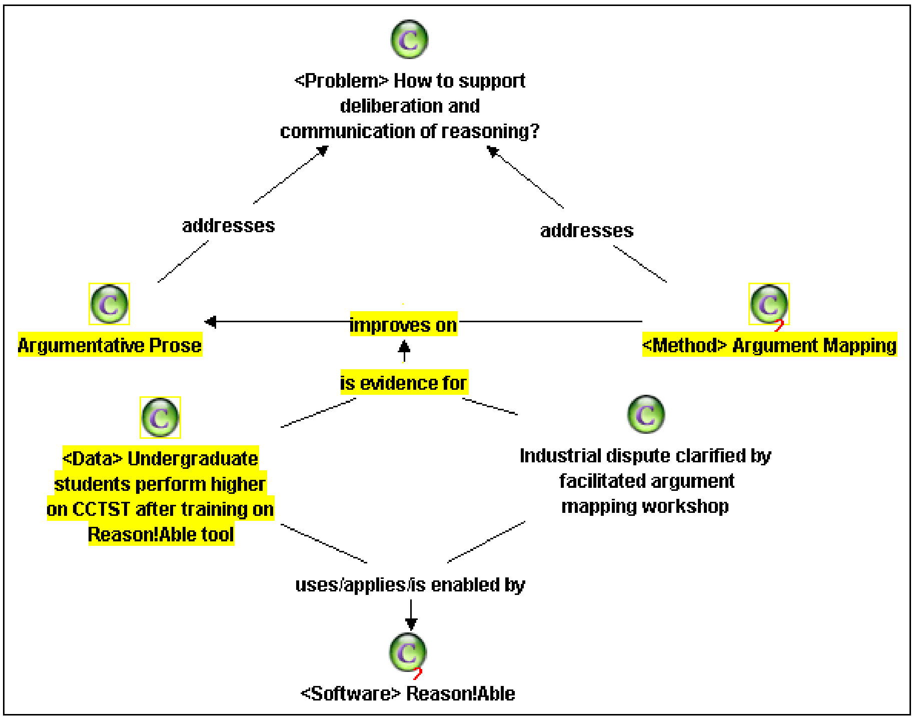

Figure 9.5: Visual claims analysis of part of Chapter 6 by van Gelder.

The map of van Gelder's Chapter 6 (Figure 9.5) illustrates not only individual claims (concept-link-concept), but claim-link-claim structures (highlighted), in other words, arguments supporting other arguments.

8.5.1 Representational Issues in Claim-Making

In any argument mapping approach (or for that matter, any conceptual modelling approach) there are always representational decisions to make about naming, classifying and linking objects, managing the coherence of the overall structure, and working at an appropriate level of granularity. Being able to make such 'meta-decisions' is a critical skill for real time collaborative argument/issue mapping approaches (van Gelder, Chapter 6, Selvin, Chapter 7), given the time and group pressures to maintain momentum; speed of capture is a major driver towards 'lightweight' notations such as IBIS. The pressure is less intense in a use context such as ClaiMaker, in which a researcher/student/analyst is working in a more reflective mode, probably (although not necessarily) on their own, distilling the essence of a piece of work into a succinct map. We can, therefore, afford a richer notation offering more expressive choices. A persistent design concern, however, is to walk the tightrope between overwhelming the user with subtly different link (and optional node) types that they cannot differentiate, and straitjacketing them into a frustratingly small vocabulary in which they cannot express themselves.

Turning to visual claim-making specifically, the use of an open, networked environment with 'live' concepts that may be used by numerous researchers (as opposed to static concept mapping in a closed application), places a premium on the *re-use* of concepts and claims wherever possible: the same idea should be expressed in the same way, as far as possible. This should be relatively simple for 'concrete concepts' such as the names of specific theories, methods, algorithms, software tools, and so forth. A quick search should reveal the concept if it has been created, so the user can just re-use it. In contrast, complex ideas will comprise claims or sets of claims of an unpredictable structure (for instance, it may impossible to know in advance how an idea such as *the internet is forcing the publishing industry to reinvent itself* will be expressed). A keyword search may reveal a good candidate for re-use, or a researcher's own knowledge of the field may take them to a document they know, whose claims they can inspect and re-use or adapt. A research group may publish a public library of concepts and claims representing their major publications, recommending that to cite their work, others should use this library (in the process, greatly assisting automated analysis of that group's research impact). These are scenarios that we can envisage, but which depend on the complex interplay of the technology and its adoption.

Another representational decision that must be made is how to lay out the structure. We can give a blank canvas for the user to lay out their arguments as they please, or provide a library of templates to 'fill in the blanks' for canonical genres of paper in a given field. This may help beginners given the evidence from educational concept mapping that some 'scaffolding' can be helpful (Reader and Hammond, 1992). Genres in the field of *Human-computer Interaction* would include *system description paper, evaluation paper* (e.g. Figure 9.6)*, theoretical paper,* and *literature review* (reflected in the categories of submission that major HCI conferences often call for). These papers have different structures, and are refereed according to different criteria; indeed it has been suggested to us more than once that authors could be required to complete a template such as Figure 9.6 to accompany their submission, also assisting referees. (The use of spider diagrams to teach students different genres of writing is of course a long established technique related to this idea of templates for conceptual, discourse-oriented publishing.)

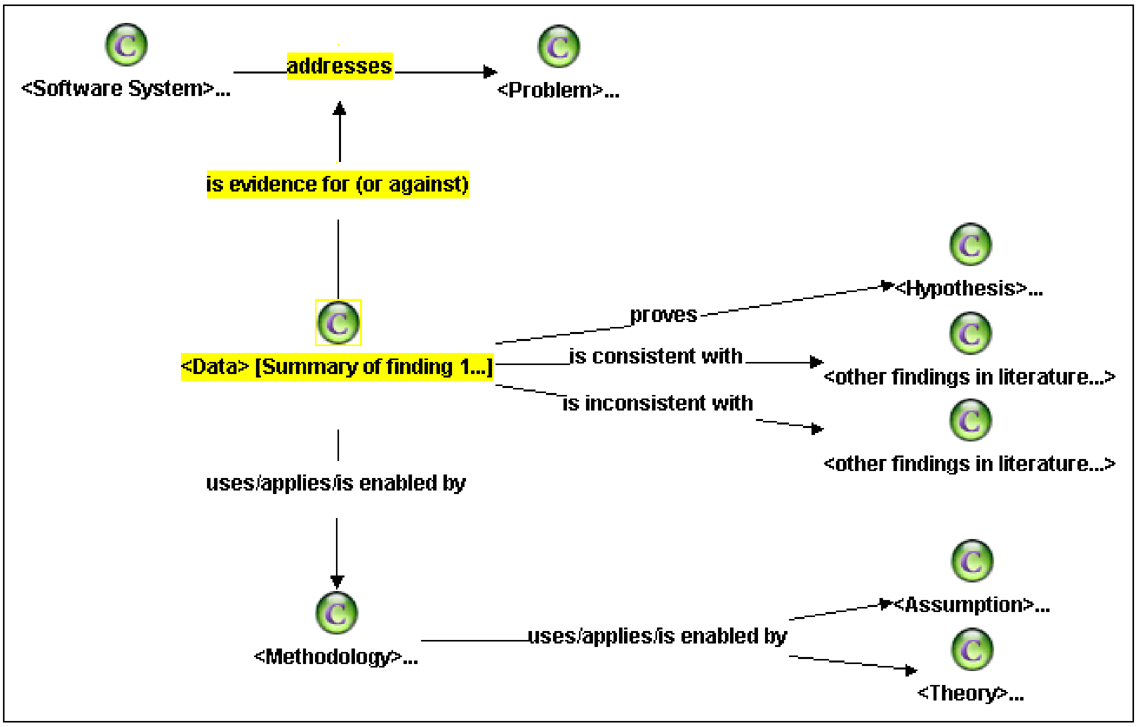

Figure 9.6: A claim-making template for a stereotypical *empirical software evaluation* paper. The structure provides scaffolding for authors to think about their work, and perhaps for reviewers to evaluate it by making it easier to trace concepts with which they are less familiar. The highlighted structure shows the expected core contribution of an evaluation paper: evidence about the effectiveness of a software system.

9.6 Analysis and Visualization of Claims Networks

Thus far, we have considered the construction of claims to populate the repository. As the network of claims and arguments grows, however, support will clearly be needed to manage the complexity. The discourse ontology has been designed to provide a language for filtering and querying. We begin this section by looking at the use of non-semantic graph theory to provide coherent views onto the network, and then illustrate the additional power gained from working from the semantics of the structure.

9.6.1 Graph Theoretic Analysis of Claims Networks

Graph theory offers mechanisms for exploring the topography of networks. In the

ClaiMaker repository, the structure that grows as claims are made can be viewed as a graph with the concepts providing vertices and the relations providing edges. We are beginning to apply techniques based on graph theory to see what phenomena of interest they can detect in a claims network. Studies on random graphs (Erdos, 1960) suggest that if you have more than half as many edges as vertices a giant component will emerge. This is a connected piece of graph that includes most of the vertices. For instance, at one point, the claims made in our early trials comprised almost as many links (531) as concepts (556), making it likely that there was a giant component. Additionally, it is possible that any giant component will be an example of a 'small world' network (Watts, 1999), which are relatively sparse (they have few edges) and are clustered. As a result they have small diameter; if directionality is ignored, a user can reach most nodes from most other nodes in a few steps (provided they know the right route). Identifying and highlighting such 'short cut' routes could play an important role in a visual browsing interface.

We hypothesise that in ClaiMaker there may be concepts that are sufficiently important that they will be used by several disciplines. For example, the concept *Small Worlds* might be linked to analysis of telecommunications networks, graph theory, and to studies of food webs. Starting a browsing session at *Small Worlds* would be helpful to a user, who could move quickly to several different regions of the graph. A first step to finding short cuts across the graph is therefore to identify clusters of highly linked documents. One way to do so is to browse filtered views of the network visually, as illustrated in Figure 9.7. Using established graph layout algorithms, augmented by interface technologies such as hyperbolic trees browsers (e.g. Inxight, 2002), one may be able to visually spot 'hub' concepts with above average numbers of links to and from them, suggesting an important concept.

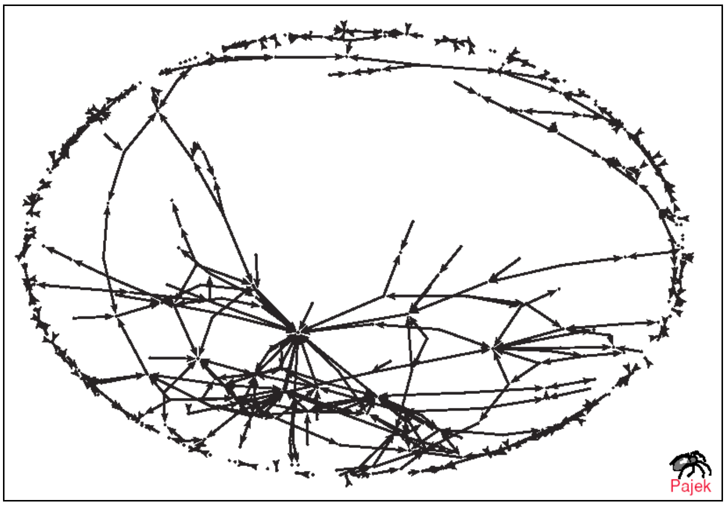

Figure 9.7: A birds-eye view of a large claim network gives a sense of gross structure such as visual cues to dense clusters of potential interest, but conveys little semantic information[4].

[4] Network visualization using the Pajek (2002) program for large network analysis.

Figure 9.8 'zooms in' to show a subgraph of Figure 9.7, filtered to show only 'significant' concept nodes (defined as having three or more claim links).

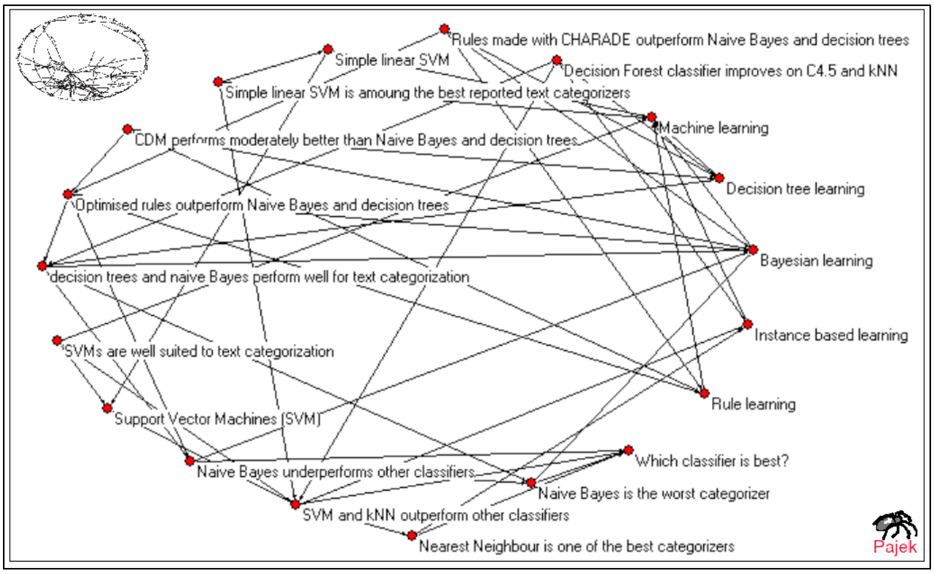

Figure 9.8: A 3-core cluster extracted from a network of claims and argumentation links. From hundreds of nodes modelling literature on text categorization, only those which connect to at least 3 other nodes in the cluster are presented (with link labels switched off). A flavour of key issues in the field is given without overwhelming the viewer.

Another approach we are experimenting with is to adapt a method from scientometrics, the quantitative study of publication and citation patterns. Scientometrics uses *citations to articles* as its basic unit of measurement, to derive 'performance indicators' of journals, individual authors, organisations, and national research efforts, or to analyse literatures for potentially significant patterns. ClaiMaker presents an opportunity to do similar analysis but at a *finer granularity*: we substitute papers with optionally typed concepts, and citations with typed relations. In doing this, we draw on work presented which demonstrates a method for treating citation networks as partially ordered graphs (Egghe, 1990). The method used for discovering highly inter-linked clusters was based on the Research Fronts method used at the Institute for Scientific Information (Garfield, 1994). This approach assumes that an interesting topic is marked by a cluster of highly cited papers, which in turn cite each other. A prototype clustering algorithm has been tested and does identify coherent topics.

The work summarised in this section encourages us to believe that graph theory and scientometrics are two of a palette of potential methods for exploring the topography of a claims network (taking no account of the semantics of the nodes and links). We turn now to analytical services which exploit the vocabulary of the discourse ontology.

9.6.2 Semantic Analysis of Claims Networks

Example 1: Perspective Analysis ("What arguments are there against this paper?")

Consider a common question that many researchers bring to a literature: *"What arguments are there against this paper?"* Despite the centrality of such a notion, there is not even a language in which to articulate such a query to a library catalogue system, because there are no indexing schemes with a model (ontology) of the world of scholarly discourse. There is no way to express the basic idea that *researchers disagree*. If we can improve on this, then we have a good example of the argumentation ontology adding value over existing retrieval methods.

How can we realise such a query? First, we are looking for *arguments against*, which map to the ontology as negative relations of any type (recall that all relations have positive or negative polarity). At a trivial level, *this paper* corresponds to the currently selected document in ClaiMaker[5]. 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*.

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

- finds the concepts associated with that paper;
- extends the set of concepts by adding positively linked concepts from other papers;
- returns claims against this extended concept set.

Typical results are presented in Figure 9.9.

5 If not already in the database (e.g. we are working with journal publishers), one can manually enter document metadata, or more conveniently, upload one's personal library of bibliographic metadata in a standard format such as Refer or Bib.

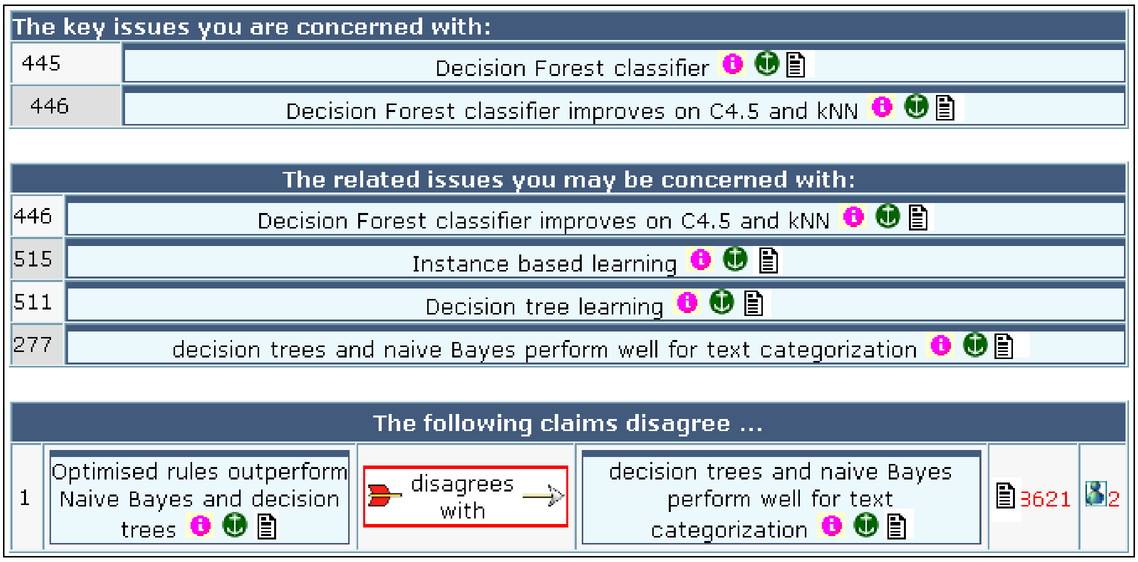

Figure 9.9: Arguments that contrast with the concepts in a research paper by Chen and Ho (2000). Key: clicking displays concept metadata; sets the concept as the focal concept, to show incoming and outgoing relations; links to the document metadata/URL. links to information about the concept's creator.

ClaiMaker then supports further structured browsing, for instance, having discovered that one of the concepts related to the article is challenged by *Optimized rules outperform Naïve Bayes and decision trees*, clicking on the icon sets this as the focal concept of interest, showing its immediate neighbourhood (Figure 9.10).

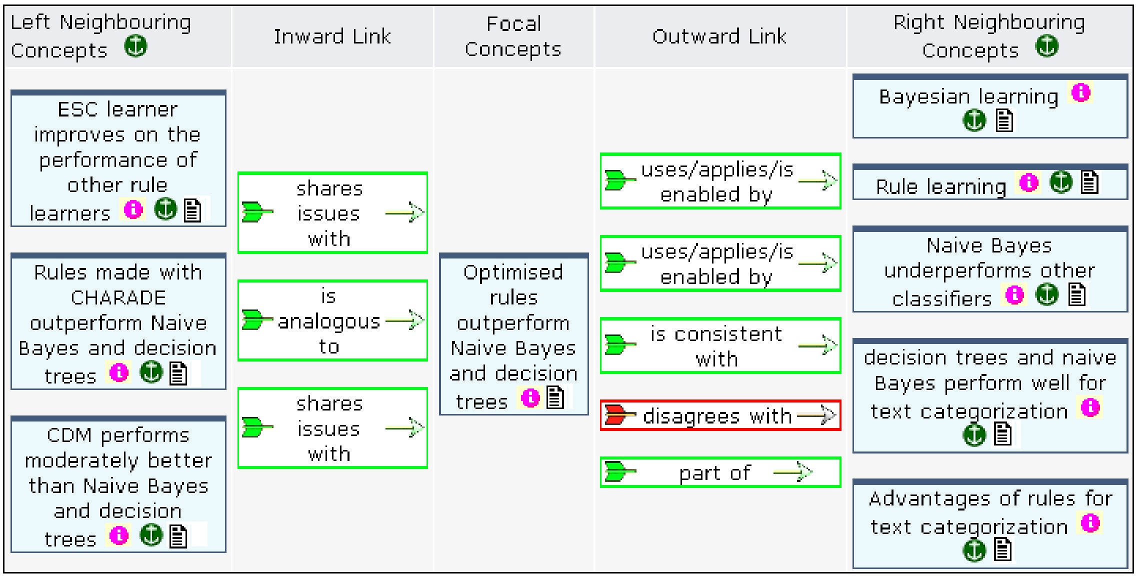

Figure 9.10: Examining the 'relational neighbourhood' around a focal concept. The concept *Optimized rules outperform Naïve Bayes and decision trees* (discovered in Figure 9.9) now occupies the centre in order to show incoming and outgoing links. Any concept displayed can then be made the focal concept by clicking on its icon.

Example 2: Lineage Analysis ("Where Did This Idea Come From?")

A common activity in research is clarifying the lineage behind an idea. Lineage is essentially ancestry and (with its inverse, the descendant) focuses on the notion that ideas build on each other. Where the paths have faded over time or been confused, uncovering unexpected or surprising lineage is of course 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. Our lineage tool tracks back (semantically, not in time) from a concept to see how it evolved, whereas 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.

So, let us consider a new query: *Where did this idea come from?* We have already suggested that a claims network can be treated as a graph, with concepts as vertices, and the links between concepts as edges. A path in a graph is a sequence of connected edges. A lineage can be conceptualised as a path in which the links suggest development or improvement. The problem of finding lineage in ClaiMaker can then be formulated as a path matching problem, a well known problem in graph theory for which algorithms exist[6].

To provide lineage analysis as a ClaiMaker service, path queries are constructed from link-types using a set of primitives. For example, we can search for paths that may be of any length, and which contain (in any order) any of the positive links that have type *similarity* in either direction, or the two general links *uses/applies/is enabled by* or *improves on*, going in the direction away from the target concept of the query. The *improves on* link type is included to reflect the notion of progress implicit in lineage, while *uses/applies/is enabled by* has a weaker implication of "building upon." The *similarity* links are included because if a new concept is like another that *improves on* a third, then the new concept may well also be an improvement. *Similarity* links are acceptable in either direction because *similarity* is a naturally symmetrical relation (if A is like B, then B is like A). Figure 9.11 shows examples of acceptable paths that could be returned by this lineage analysis.

The search can be tightened by filtering the paths returned to ensure they contain the *improves on* relation, after which only the second of the paths in Figure 9.11 would be retained. Conversely, one can relax the conditions to broaden the search, for instance, to permit the inclusion of any Problem-related links (see Table 9.2), since *addressing* or *solving* a known problem usually represents progress of some sort. One could also include Taxonomic links, since if a *part of* some innovation *improves on* another approach then it implies there may be improvement overall. Note that in these cases, the direction of the link is fundamental: it is only problems that the new concept *solves* that are of interest, and even if a whole innovation is an improvement, there is no reason to assume that every *part of* it is also. One advantage of the path matching approach is that

---

6 A semantic web standard based on graphs is the *Resource Description Framework* <www.w3.org/RDF>. In the analysis presented here we use the *Ivanhoe* path matching tool available in the *Wilbur* RDF toolkit <wilbur-rdf.sourceforge.net>.

it facilitates the use of directional elements in queries.

The results of this kind of structural query can then be rendered in a variety of forms back to the user. Figure 9.12 shows a visualization of the structure extracted from the claims network in response to a lineage query about a concept.

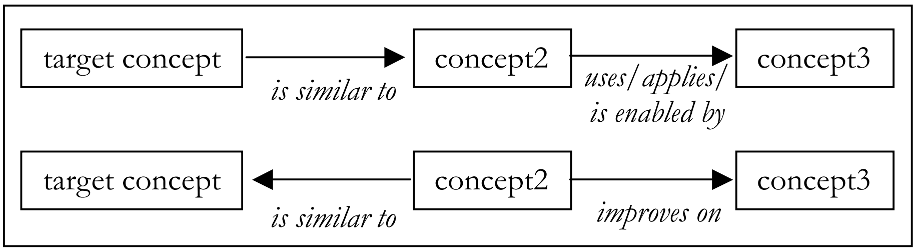

Figure 9.11: Examples of paths that could be returned by a lineage analysis on a target concept (see text for the specification of the query).

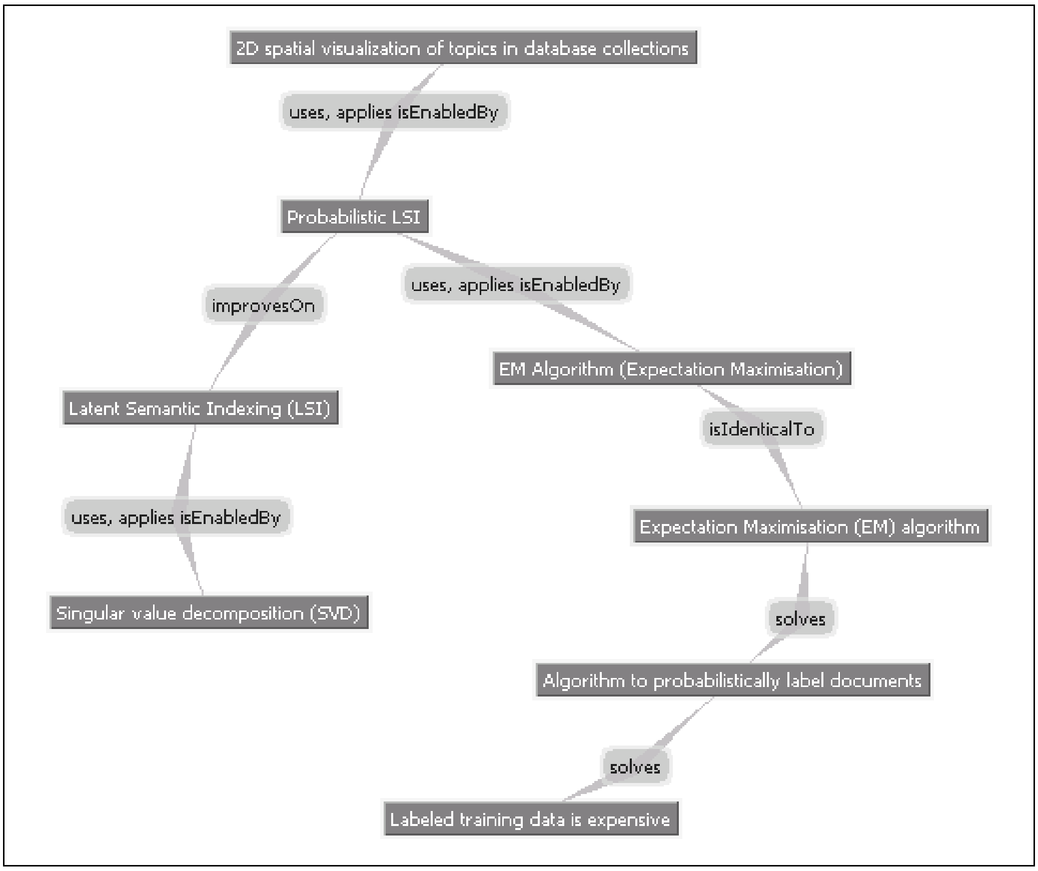

Figure 9.12: Visualization of the results of a *lineage analysis*, a representation of the claims in the network on which the top concept explicitly and implicitly builds, or alternatively, a guide to the local context in which a concept is embedded[7].

7 Graph visualization courtesy the *Ceryle Project* by Murray Altheim, Knowledge Media Institute, Open University <kmi.open.ac.uk/projects/ceryle/>. Ceryle includes an enhancement of the *TouchGraph* graph

The lineage function (and its inverse, descendants) can be thought of as providing an analytical tool to excavate the *foundation* under an idea (or conversely, an indicator of its *impact*). From a navigational perspective, they can be thought of as offering *focused browsing tools.* In response to a *"Where am I?"* question, they give answers in terms of developmental context, positioning ideas in the literature in terms of their evolution.

To summarise, term-based information retrieval handles documents as isolated entities defined by the words in them. Citations in a document give no indication of authors' intentions in referring to other work; we cannot even tell if a paper is referenced because the authors support or are diametrically opposed to it. The examples of *Perspective Analysis* and *Lineage Analysis* demonstrate how the discourse ontology can make the connections between ideas in different documents explicit, enabling novel, powerful kinds of query.

9.7 Conclusion

We have identified a striking absence of tools for global research argumentation. As the tidal wave of online information builds, the ability to model and analyse expert debate and evidence on research problems will grow in importance. The Scholarly Ontologies project is envisioning how research publishing and discourse could evolve over the next decade, given current infrastructure developments. This is self-evidently a large scale project, but one which we hypothesise to be both socially and technically tractable. As a 'wicked problem' (Rittel, 1984; Buckingham Shum-Chapter 1, van Bruggen-Chapter 2), we have to build realistic tools to understand the problem space, and so are implementing the ClaiMaker system to mediate structured, distributed argumentation.
We have summarised the current status of this work in progress: a discourse ontology with a specific focus on scholarly argumentation moves, associated prototype analysis tools to assist in managing the complexity of a collaboratively built semantic graph of claims and counter-claims, and we have described the cognitive design issues that are arising in the creation of user interfaces for 'visual claim-making' with 'live' concepts and links that many other researchers may be using.

We are releasing versions for interested members of any research community to start modelling their literatures, for instance, to assist research, teaching or information analysis. The only way that a new infrastructure grows is when individuals recognise the value that it can add to their work. As significant examples grow, we aim to demonstrate the value of various information services for managing the complexity (e.g. visualization for teaching; structural querying; alerting services). We welcome approaches from colleagues who wish to be early adopters, and join us in mapping the new territory that is opening up.

---

visualization toolkit by Alex Shapiro <touchgraph.sourceforge.net>. TouchGraph delivers interactive, self-organizing maps over the web via a Java applet.

9.8 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. Special thanks also to Ora Lassila for expert advice on *Ivanhoe* for lineage analysis, and to Murray Altheim for use of *Ceryle* for lineage visualizations.

9.9 References

Boland, R. J. J., & Tenkasi, R. V. (1995). Perspective making and perspective taking in communities of knowing. *Organization Science*, *6*(4), 350-372.

Buckingham Shum, S., MacLean, A., Bellotti, V., & Hammond, N. (1997). Graphical argumentation and design cognition. *Human-Computer Interaction*, *12*(3), 267-300.

Buckingham Shum, S., Motta, E., & Domingue, J. (2000). ScholOnto: An ontology-based digital library server for research documents and discourse. *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. Invited contribution to: *National Science Foundation Workshop on Infrastructures for Distributed Communities of Practice*, San Diego, CA. Retrieved on August 1, 2002 from http://kmi.open.ac.uk/projects/scholonto/docs/SBS_DCP2002.pdf

Chen, H., & Ho, T. K. (2000). Evaluation of decision forests on text categorization. *Proc. 7th SPIE Conference on Document Recognition and Retrieval*, 191-199.

Garfield, E. (1994, October 10). Research fronts. *Current Contents.* Retrieved on August 1, 2002 from http://www.isinet.com/isi/hot/essays/citationanalysis/11.html

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, *43*(5/6), 907-928.

Knott, A., & Mellish, C. (1996). A feature-based account of relations signalled by sentence and clause connectives. *Language and Speech, 39*(2-3), 143-183.

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.

Egghe, L., & Rousseau, R. (1990). *Introduction to informetrics: quantitative methods in library, documentation and information science*. Amsterdam: Elsevier.

Egghe, L., & Rousseau, R. (in press). Co-citation, bibliographic coupling and a characterization of lattice citation networks. *Scientometrics.*

Inxight *Star Tree*. Retrieved on August 1, 2001 from http://www.inxight.com/products/core/star_tree

Li, G., Uren, V., Motta, E., Buckingham Shum, S., & Domingue, J. (2002). ClaiMaker: weaving a semantic web of research papers. *1st International Semantic Web Conference*, (Sardinia, June 9-12th, 2002). Retrieved on August 1, 2002 from http://kmi.open.ac.uk/projects/scholonto/docs/ClaiMaker-ISWC2002.pdf

Mancini, C., & Buckingham Shum, S. (2001). Cognitive coherence relations and hypertext: From cinematic patterns to scholarly discourse. *Proc. ACM Hypertext*

2001, (Aug. 14-18, Århrus, Denmark), 165-174. New York:. ACM Press Retrieved from http://kmi.open.ac.uk/tr/papers/kmi-tr-110.pdf

Newman, S., & Marshall, C. (1991). *Pushing Toulmin too far: Learning from an argument representation scheme* (Technical Report SSL 92-45). Xerox Palo Alto Research Center.

Erdos, P. A. R. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences, 5*, 17-61.

Pajek (2002). *Program for Large Network Analysi*. Retrieved on August 1, 2002 from http://vlado.fmf.uni-lj.si/pub/networks/pajek/default.htm

Reader, W., & Hammond, N. (1994). Computer-Based Tools to Support Learning from Hypertext: Concept Mapping Tools and Beyond. *Computers in Education*, 22, 99-106.

Rittel, H. W. J. , & Webber, M. M. (1984). Planning Problems are Wicked Problems. In N. Cross (Ed.), *Developments in Design Methodology* (pp. 135-144). Chichester: John Wiley & Sons. (Published earlier as part of "Dilemmas in a general theory of planning", *Policy Sciences*, 4, 155-169, 1973).

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 Cooperative Work*, *8*(4), 333-352.

Toulmin, S. (1958). *The Uses of Argument*. Cambridge: Cambridge University Press.

Trigg, R., & Weiser, M. (1983). TEXTNET: A Network-Based Approach to Text Handling. *ACM Transactions on Office Information Systems*, *4*(1), (pp 97-100)

Watts, D. J. (1999). *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton, NJ: Princeton University Press.