Invited Discussion Paper
2nd International Workshop on Infrastructures for Distributed Collective Practices
San Diego, 6-9 Feb. 2002

Designing Representational Coherence into an Infrastructure for Collective Sensemaking

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

Knowledge Media Institute, The Open University, UK

Contact: sbs@acm.org
www.kmi.open.ac.uk/projects/scholonto

Abstract: We discuss issues arising from the design, implementation and first use of a prototype infrastructure for distributed collective practice (IDCP), and reflect upon their intersection with some of the themes emerging from the Paris 2000 IDCP workshop. The problem of maintaining *coherence* in a distributed system is of central interest to us. We focus on the notion of *representational coherence*, and consider both *process* issues (the evolution of a discourse structuring scheme; tracing infrastructural history), and the affordances of the resulting *product* (uncertainty with respect to the scheme's application; ways to map the topography of the emergent representation, with particular interest in graph theory). Throughout, we highlight issues that could have broader implications for IDCPs.

Keywords: representational coherence, rhetorical relations, interpretation, digital libraries

1 "Coherence"

Coherence is one way to encapsulate some of the dominant themes that characterise the emerging DCP research agenda. Words are just words, so what does this one buy us? If something is "coherent", it "makes sense". It evokes a sense of wholeness in the perceiver. Coherence is not tied to textual forms, or specific logic systems. Visual and performing arts, cinema, and dance have all evolved complex signification systems for expressing coherence, interweaving both rational and emotional dimensions. A mathematical equation and a scientific visualization have coherence of different sorts. Coherence comes from an underlying abstract discipline (reflected in the critical language spoken by experts) which defines the form. Coherence need not come at the cost of ironing out ambiguity, heterogeneity, or inconsistency; indeed, it may depend on them to provide the textures and contrasts out of which emerge the more complex whole. Coherence thus exists at different levels and granularities of description. Coherence, like beauty, is in the eye of the beholder, and assumes a level of literacy to translate the stimuli into meaning (hence the inscrutability of mathematics, opera, literature, and art to different groups). Coherence is rooted fundamentally in one's perspective: *what* one expects to see, is *able* to see, and *why* one wants to see.

The preceding list can be taken as evidence that *coherence* is a vague (albeit rich) concept that gives us little purchase on important theoretical or design problems. Conversely, it may be read as a way to frame IDCP research precisely because of this richness (a similar line is taken in related research into *narrative* as a ubiquitous, fundamental mode of cognition). IDCP research needs ways to talk about *formal coherence* as implemented in metadata, interoperability standards, ontologies and artificial reasoning, as well as the tacit, situated, contingent sensemaking in which individuals and groups engage at the workplace, and specifically, when they are part of a distributed group. Moreover, there is promising theoretical and computational work that provides systematic, implementable measures of coherence that could be applied to IDCPs, some of which we will touch on later.

A community of practice (CoP) establishes and sustains forms of coherence that order their activity, and by definition, creates boundaries. So, in a mature IDCP one would expect to find *coherence in practices* that demonstrates an understanding of how to choreograph, or orchestrate (choose your favourite coherence metaphor), collective activity within the constraints of the hardware, software and material infrastructure. More specifically, a key element of a CoP is its conventions in codification, that is, in its construction of symbolic artefacts. This is part of what we shall call *representational coherence*. One concern here is the way in which digital resources (research *Claims* in the case of ScholOnto – see next section) are codified at the *structural* level (chunked, classified, linked), and the consequences for detecting significant patterns in the resulting network that correspond to coherence at the *semantic* level (i.e. the coherence of the ideas which the network mirrors). Another concern is the extent to which the rationale underpinning infrastructural decisions can be 'excavated'.

Codification schemes, including taxonomies, metadata schemes and ontologies are, of course, double-edged swords. If imposed, they can straitjacket and alienate people, but through participatory negotiation of their design, they can build common ground between stakeholders. They can reduce cognitive overheads by filtering information, or add cognitive overheads by imposing an arcane and rigid grid at the wrong granularity, obstructing rather than assisting resource discovery or encoding. They can elegantly encapsulate valuable intellectual work in a widely accessible form, or bring with them suspect assumptions that are hard to uncover or undo as further schemes are layered on top.

Despite the prevailing emphasis in DCP research on divergence, heterogeneity and decentralisation, we must recognise of course that not all DCPs are anarchic collectives inventing new local dialects each week! The members of a DCP may subscribe rigorously to a classification scheme. But a 'DCP claim' seems to be that, regardless of how disciplined a DCP is, distributing individuals in time, space, and background inevitably aggravates "ontological drift" (to borrow from Robinson and Bannon [10]) from agreed convention.[1] Case studies are needed to characterise in what contexts this is particularly likely, and

---

[1] Bannon and Robinson focused on the role of disciplinary heterogeneity in changing the interpretation of design specifications. In distributed settings even without such differences, another cause of 'drift' may be simply that opportunities for realignment when differences emerge are likely to be reduced as a result of fewer, and lower bandwidth, encounters. Developers of high bandwidth and/or 'lightweight/ opportunistic' collaboration media would certainly argue this.

perhaps, if there are contexts in which distributed members work more effectively than comparable co-located communities in coordinating themselves around a formalism.[2]

To cover the whole lifecycle of a classification scheme, a 'formalism-aware' IDCP would, ideally, support its members in:

- negotiating the design of a new scheme, or evolution of an existing one
- capturing the rationale behind decisions
- releasing the scheme for use
- tracking its usage
- negotiating revisions, and maintaining/upgrading older versions

We touch on all of these briefly in this paper.

2 Overview of ScholOnto

There remains a yawning gap in the researcher's digital toolkit: as digital libraries add by the day to the ocean of information, there remain few tools to track ideas and results in a field, and to express and analyse positions on their *significance*. It is not surprising that such tools should lag behind the publishing of datasets/information online, since talking about *interpretation* (and hence *meaning*) is a rather harder problem. Nonetheless, researchers require 'sensemaking' tools. Written prose remains the way in which interpretations are expressed, but ironically, publishing new documents to make sense of existing documents also adds to the information glut. Citation analysis tools are a start, but are relatively blunt instruments for post hoc analysis. At present, freeform annotation and discussions are the only online forums for researchers to *express* interpretations, but these generally have little formal structure, and have little status in scientific publishing, so consequently are perceived as very informal media.

The *Scholarly Ontologies* (ScholOnto) project [1] is exploring an alternative scenario. We are developing an ontology-based 'Claims Server' to support scholarly interpretation and discourse, investigating the practicality of publishing not only documents, but associated conceptual structures in a collective knowledge base. The system enables researchers to *make claims*: to describe and debate, in a network-centric way, their view of a document's contributions and relationship to the literature. It thus provides an *interpretational layer* above raw DLs. This contrasts with most DL/semantic web applications that require consensus on the structure of a domain, and an agreed metadata scheme that tries to iron out inconsistency, ambiguity and incompleteness. ScholOnto is all about supporting

---

[2] DCPs advocating a standard (such as Dublin Core or Topic Maps) are arguably amongst the most tightly regimented of DCPs in their adherence to common formalisms, but they are for this reason also atypical: their raison d'être is to prove that a DCP *can* coordinate itself around *their* artifact, and thus demonstrate its usability, scalabilty, and interoperability. Would a co-located community do this less effectively? One might think so, but only because the lack of "D" in their "CP" would make their collective practice weaker at detecting problems in a formalism whose primary design requirement is that it can work in a distributed environment.

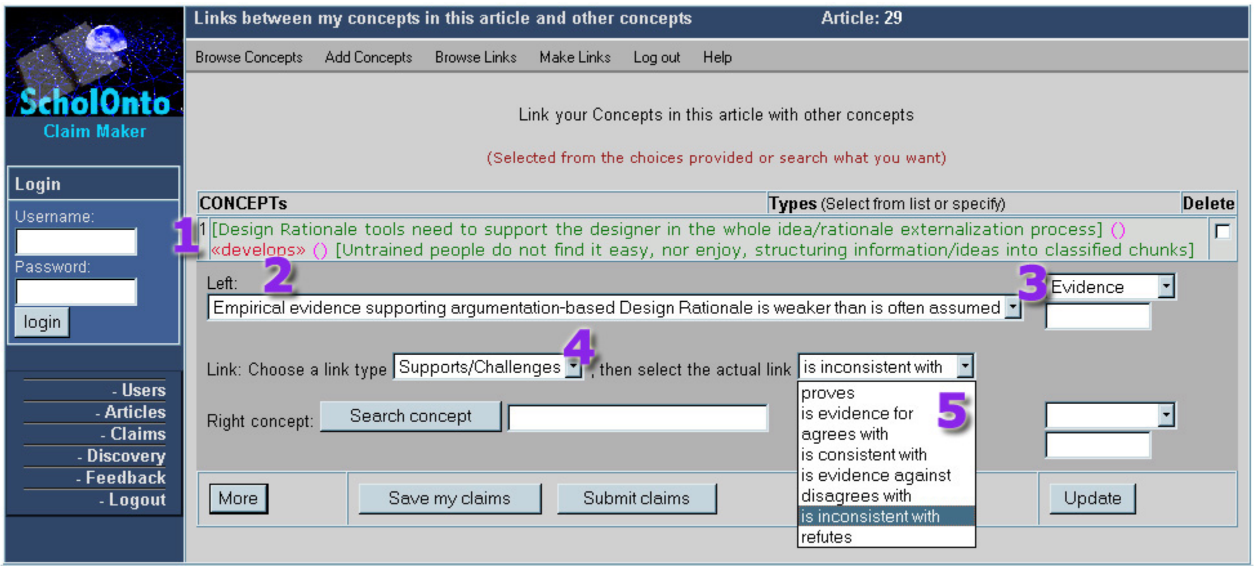

Figure 1: User interface to the ScholOnto Claim Server, showing how a researcher can build a set of claims. Key: [1] A claim that has already been constructed, ready to submit; **[2]** a concept to link *from*, **[3]** assigned type *Evidence*, and **[4]** linked via the Relational Class *Supports/Challenges*, **[5]** more specifically, *is inconsistent with* (selected from the dialect-specific menu).

principled disagreement, conflicting perspectives, and the resulting ambiguities and inconsistencies, because they are the very stuff of research, and the objects of explicit inquiry. A preliminary discussion of DCP-related issues was presented at the Paris 2000 workshop [2], and a detailed rationale and discussion of other HCI/CSCW related issues is given in [1].

Figure 1 shows a prototype interface for researchers to make Claims about documents.

We now give some tangible examples of IDCP phenomena and design challenges as they manifest in ScholOnto. Throughout, we weave discussion of possible implications for IDCPs more broadly, considering the extent to which elements of ScholOnto's approach might be generalised.

3 Defining and evolving a discourse structuring scheme

In this section, we present a brief case study of the design, deployment and revision of a ScholOnto's scheme for structuring scholarly discourse as Claims. We present this process-oriented description as an example that may be of interest to others developing IDCPs that seek to constrain interactions, and it also sets the background for subsequent discussion of how the concept of "coherence" takes form in the system.

The ScholOnto ontology is novel, requiring a scheme led by relations rather than concepts. Consequently, we took an evolutionary design approach. A prototype scheme was devised, and an input system built around it and used for a few months by the research team. The claims we created provided a body of data, which was reviewed. The review process assessed the representational coherence of our prototype scheme, tackling inconsistencies and sources of ambiguous claims.

A list of requirements for the first ontology was drawn up by examining both theoretical issues, accessed through a literature search, and practical ones, brought out by producing "mind maps" of a selection of microliteratures. Table 1 summarises these requirements. An issue for discussion is whether they have broader application to IDCPs.

Requirements for a research 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.** A classical logic model grounded on notions of "truth" would not fit our purpose; if "truth" is established on an issue, it ceases to be worth doing research about. Our scheme must therefore be closer to that presented by Toulmin [12], with evidence being presented in favour of claims and complemented by counter-claims. To support or challenge a claim, one 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 [4,5] (summarised and related in detail to ScholOnto in [7]). If you agree with a proposition, the relation used would have positive polarity. If you 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.
3. **Ownership of public content is critical.** Contributors must take responsibility for the claims they make. ScholOnto'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 ScholOnto 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 [12].
4. **Social dimensions to being explicit.** ScholOnto invites researchers to consider making explicit what is normally implicit in the text of a paper (discussed in [1]). Discourse relational types vary in strength, which has both computational and social dimensions. 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. **A concept has no category outside of use.** Concepts may or may not be typed and 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. 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 several types in different situations.

6. **The scheme should 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 on Cognitive Coherence Relations [4,5], 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. *objects to*), but the basic properties of the strongly negative relation that challenges a concept would remain unchanged. This method would let us configure ScholOnto for different communities without altering the underlying reasoning engine.

Table 1: Requirements for a research discourse ontology

3.1 Ontology v1

The first version of the 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 was a set of metadata describing a document in a digital library, these provided the backing, every claim being grounded in a paper. Concepts were stored as short pieces of free text. A claim was a triple of two objects linked by a relation (Figure 2). Each relation had the properties type, polarity and weight, and a dialect label in natural language. A concept may optionally have one or more concept types, each stored as part of a claim. By storing the concept type in the claim, rather than attaching it permanently to the concept, the typing of concepts could be made context dependent.

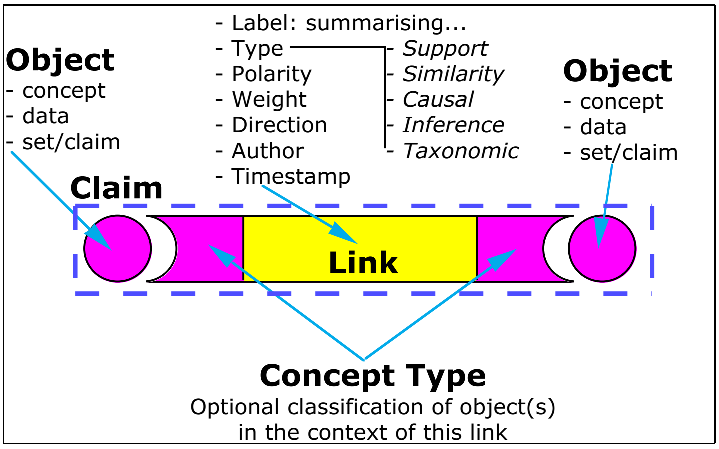

Figure 2: Structure of a *Claim* in the first version of the ontology

Five rhetorical types were chosen for the links based on the literature review and the modelling exercises. These were Supports, Similarity, Causal, Inference and Taxonomic. A dialect was written for each of the first four types that offered a spectrum of links from strong positive to strong negative. For example, the dialect for similarity had eleven relations, which ranged from "is identical to" (positive polarity, weight 5), to "is the opposite of" (negative polarity, weight 5). Our aim was to offer a wide choice and see which links we used. For the Taxonomic class, we offered the standard relations instance of, part of and subclass of.

3.2 Revising a discourse scheme and updating the knowledge-base

One of the generic problems with any ontology or database schema is updating the data/knowledgebase when the schema changes. The problems are obviously aggravated if numerous data/knowledge bases based on the schema exist, and infrastructures to assist with distributed management are a research problem in their own right. The v1—v2 revision described here was on a central server, avoiding this complication. We summarise this process before drawing some conclusions.

Once more than 500 claims had been made using the first version of the ontology, the entire team sat down to review the claims, and identify weaknesses in the scheme (captured in Compendium). We looked for usability problems, and for signs of incoherence.

One of our aims in the review was to reduce the number of link labels available. The first set of dialects had deliberately offered a large selection, in order that we could find out which labels people would use. However, it had proven hard to choose from so many similar options. It was felt that a smaller selection of labels with greater semantic differences between them would make the scheme less ambiguous. The new version drastically pruned the number of dialect labels from 62 to 32.

We found it necessary to introduce a new *General* relational class, whose members were drawn from several other classes. This class incorporated a group of neutrally weighted links, some of which had turned out to be the most heavily used links in the original dialect, including *uses* (128 incidences), and *concerns* (67).

We had also identified the need for a link type that referred specifically to concepts typed as *Problems*. Finding solutions to problems is clearly a key research activity, and the concept type *Problem* had been regularly used. Finally, we eliminated the *Inference* relational class, finding several ambiguities with *Supports/Challenges* and *Causal* dialects.

Some changes proved simple, e.g. we changed the label of the *instance of* link to *example of* to make it more generally understood. Others sparked impassioned debate: is *has nothing to do with* a valid relation? Why was *has something in common with* been so frequently (and sloppily) used? What is the opposite of *improves on*?

The revised version of the ontology is shown in Table 2. Some points worth making are that the weighting scheme has been radically simplified, there are now effectively two weights, 2 and 1, implying strong and weak relations. The inference type has disappeared and been absorbed into *Supports/Challenges*. Two new types *Problem Related* and *General* have emerged. A mapping script converted the original claims to the new scheme, and we have confirmed that the revised version covers all the ground that the original version did. Having tightened up the distinctions and eliminated unhelpful ambiguities, we have greater confidence that the new relations have the representational coherence required to encourage contributors to make clean, easily understood claims. This is about to be tested in a larger, but still small scale release, followed by a public release.

RelationClass	Dialect label	Polarity/Weight
General	is about	+/1
	uses/applies/is enabled by	+/1
	improves on	+/2
	impairs	-/2
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
Taxonomic	part of	+/1
	example of	+/1
	subclass of	+/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 analagous to	+/1
	is not analagous to	-/1
Causal	predicts	+/1
	envisages	+/1
	causes	+/2
	is capable of causing	+/1
	is prerequisite for	+/1
	is unlikely to affect	-/1
	prevents	-/2

Table 2: The revised discourse ontology

In sum, we confirmed three things from this exercise:

1. **Version 1 won't be right.** It is probably impossible to define a representational scheme for communication a priori: no matter how principled a communication scheme it is, or how much it draws on a wide literature and established models, it has to be deployed and used to evaluate its match to the specific communicative purposes and constraints under which one is operating.

2. **Version mapping worked.** Aware that v1 was just the 'first cut', we had hoped that mapping rules from v1 to v2 could be written. This trial confirmed that they could indeed be executed with coherent results, at least on a centralised server.

3. **The modular structure of a Claim was robust**, assisting both computationally and with the 'cognitive hygiene' of the mapping process. Polarity was a robust attribute: there were no cases where polarity reversed in v1-v2 mappings. The v1 relational classes assisted in the mapping process, since where a class was merged with another, links from that class could be mapped to the destination. Weighting proved robust, in that the new *General* class in v2 incorporated commonly used, neutrally (0) weighted links from other classes.

In a larger scale scenario, following extended use of ScholOnto by a widely distributed community, 'upgrading' the ontology would be more complex, but this exercise gives us confidence that it would be tractable. An important principle would be 'backwards compatibility', such that existing claims using old versions would not be changed irrevocably or without consultation, possibly rendering them meaningless (mapping rules always have limitations). The notion of layered version 'masks' is a possibility, whereby a link could have an original type, with subsequent masks 'on top' for later versions showing how it will be treated in computations (derived from the mapping rules). Data is never lost, but remains compatible and interoperable (to a degree) with the current version.

4 Coherence of ideas in ScholOnto

In the introduction we painted a broad picture of how we might think about coherence in relation to IDCPs. A specific kind of coherence of interest in ScholOnto relates to its particular domain of concern: the coherence of research ideas.

The first thing to clarify is that ScholOnto is neither a quality control mechanism, nor a truth maintenance system in the technical sense. Any moderation of submissions is delegated to human reviewers should the user community decide to create such a mechanism (although we anticipate the system critiquing submissions, alerting researchers to possible weaknesses). However, purely on the basis of structure, its representational scheme provides the basis for services that could assist in the analysis of "coherence".

In terms of typed concepts and relationships, we can spot 'structural signatures' that may signify interesting phenomena:

- **A common approach?** Clusters of concepts that appeal to common concepts as their theoretical basis would be a structural pattern of interest.

- **Incoherent concept?** Incidences of a concept (or author) apparently disagreeing with itself would be a structural pattern of interest; similarly, incidences of a concept (or an ancestor/descendant) apparently supporting *and* challenging the same concept (or an ancestor/descendant), or pursuing two goals that are apparently incompatible, or in tension of some sort (negative links between them)

See [1] for more examples of this sort, that we can implement in OCML as knowledge level descriptions. We discuss next a particular class of topographic structures that graph theory helps to identify.

5 Topography of distributed collective interpretation?

Graph theory offers mechanisms for exploring the topography of an IDCP, or more precisely, the phenomena associated with the infrastructure. This is well established in the generation of Social Network Analysis sociograms used to map social ties between actors [13]. In ScholOnto, the structure that grows as claims are made can be viewed as a partially ordered 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. If promising, the graph theoretic perspective can then be added to the reasoning mechanisms native to the OCML knowledge modelling environment on which we are building [8].

Studies on random graphs [9] 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, the claims made in our trial of ScholOnto v1 comprise almost as many links (531) as concepts (556). Therefore, it is very likely that there is a giant component. Additionally, it is possible that any giant component will be an example of a small world network [14]. Such networks are relatively sparse, with few edges, are clustered, and have small diameter. What this means is that, if directionality is ignored, you can reach most nodes from most other nodes in a few steps, provided you know the right route. Helping contributors to explore the claim space using visualisation methods is one of the objectives of the project. Identifying and sign posting short cut routes could play a role in a browsing service.

We hypothesise that in ScholOnto there may be concepts that are sufficiently important that they will be used by several disciplines (more broadly in IDCPs, this might correspond to different communities of practice). 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 as you 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.

As a first experiment in identifying clusters we adapted a method from scientometrics, the quantitative study of publication and citation patterns. It uses citations to articles as its basic unit of measurement and can present performance analysis of the literature about journals, individual authors, organisations, and national research efforts. ScholOnto presents an opportunity to do similar analysis at a finer granularity. In doing this, we draw on work presented which demonstrates a method for treating citation networks as partially ordered graphs [6]. The method used for discovering highly inter-linked clusters was based on the Research Fronts method used at the Institute for Scientific Information [3]. This approach assumes that an interesting topic is marked by a cluster of highly cited papers, that in turn cite each other. By analogy, we substitute papers with optionally typed concepts, and citations with typed relations. A prototype clustering algorithm was tested on the database of claims. A sample of the clusters of concepts found is given in Table 3.

Cluster 1

- Conditions for social behaviours
- Social behaviour in networked information systems

Cluster 2

- An ontology is an explicit specification of a conceptualization
- Formalism
- Minimizing problems of formalism
- Ontology driven document enrichment

Cluster 3

- Questions, Options and Criteria (QOC)
- Argumentation-based design rationale is useful
- Argumentation schemes augment reasoning
- QOC facilitates the representation of design rationale
- QOC obstructs the representation of design rationale

Table 3: Results of link based cluster analysis, showing the text label of each concept.

While this is clearly a heuristic (and even simplistic) approach, it does seem to have identified some topics of interest with relative ease. This 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, that is, a picture of the community's collective interpretation of its knowledge. To return to Social Network Analysis, an idea often suggested to us, but which we have yet to explore, is to apply SNA to a claims network in order to uncover social ties between the researchers behind a literature.

6 Coping with uncertainty over the use of a representation

'Heterogeneity' (of all sorts of things) seems to be a dominant theme in DCP research. Issues of heterogeneity in ScholOnto manifest most obviously in the way in which researchers make their claims. Researchers should be able to use ScholOnto without formal training, although, of course, practise will build confidence and fluency. There will, inevitably be variability in the way in which two researchers make similar claims, since we do not moderate submissions, terminology and disciplinary background varies, as does the purpose and context for making claims. (The same comments might apply to many other activities in a DCP.)

In an early trial of ScholOnto, two of the team analysed a small corpus of documents and constructed networks of claims around them *without looking at what the other had done* (in contravention of what will become recommended practice to encourage reuse). This test was run to put version 1 of the ontology and user interface through its paces, as well as to see the variability in claims that could result in order to analyse the possibilities for preventing undesirable duplication of the same claim. As an example of the results, below are claims about the same paper by two different users (Mike and Andy).

We see differences in the level of abstraction, for instance the claim:

> **4. Creator:** Mike **Date:** 2001-07-27 (conceptID=203 and ArticleID=29)
> (Article title: Argumentation-based design rationale: what use at what cost?)
> [Evidence] "argumentation schemes augment reasoning"

—is more general than:

> **10. Creator:** Andy **Date:** 2001-10-12 (conceptID=565 and ArticleID=29)
> (Article title: Argumentation-based design rationale: what use at what cost?)
> [] "Argumentation-based Design Rationale can assist certain forms of design reasoning"

We see differences in coding style - Mike's concepts appear abrupt by comparison with Andy's, because he has broken down the arguments into atomic units, which are then linked to build richer argumentative structures:

> "argumentation schemes augment reasoning"
> <<uses/applies/is enabled by>> "QOC helps bring out implicit assumptions"

We also see differences in which aspects of the document were made explicit:

> **1. Creator:** Mike **Date:** 2001-07-27 (conceptID=199 and ArticleID=29)
> (Article title: Argumentation-based design rationale: what use at what cost?)
> [] "Argumentation-based design rationale is useful"

—is a claim that the paper sets out to *test*, with one *conclusion* being:

> **8. Creator:** Andy **Date:** 2001-10-12 (conceptID=563 and ArticleID=29)
> (Article title: Argumentation-based design rationale: what use at what cost?)
> [] "Empirical evidence supporting argumentation-based Design Rationale is weaker than is often assumed"

This was a worst-case scenario in which contributors ignored what had already said, but this can and will occur in a distributed, unmoderated infrastructure. We are at an early stage in our trials, but when we release the system to early adopter communities (set for Spring 2002), we will study what phenomena emerge in larger scale use. The key question is whether the system is coherent enough to tolerate different claim making styles and still deliver discovery services that end-users value.

That being said, ScholOnto's notational structure can provide helpful constraints to encourage researchers to reify the same idea in the same way. The system uses some basic knowledge of "well-formedness" to prevent claims that are non-sensical. An example would be:

- [Method] X *is an example of* [Language] Y
  (examples have to be of the same type)

The mappings between many other concept types are, however, hard to constrain on purely logical grounds, providing a lot of expressive flexibility to authors. Interface design has a role to play here, helping to reduce the proliferation of undesirable duplications:

- For a given document, one can scan a list of claims that have been made concerning it. If a researcher sees the one that she wants to point to, then this saves her time and effort. There is thus a motivation to reuse rather than duplicate.
- If a researcher for whatever reason tries to 'reinvent' an existing concept under a known name, it is trivial for the system to alert him that it already exists.

7 Infrastructural transparency

Before concluding, we return to an important theme to emerge from the Paris 2000 workshop, concerned with the longer term evolution of IDCPs: infrastructural transparency. Bowker and Star have drawn attention to the opacity of classification schemes as a significant problem when trying to recover rationale and implicit assumptions behind accepted standards. Once associated infrastructure begins to be built on top of these standards (in the context of IDCPs, perhaps an XML DTD or an ontology-based web service), this class of historical information sinks 'lower' into the infrastructure, and the effort required to 'excavate the foundations' increases.

An approach to tackling this is Compendium, a hypertextual discourse structuring tool [2]. Since 2000, we have been using this to augment the design process of ScholOnto (and several other semantic web applications) with rationale and process capture. Compendium helps construct a contextual space around an artifact, capturing ideas, design options and open questions that are otherwise lost. We propose that tools like this can serve as the 'thin end of a wedge' to prise open the joints of an engineered artifact (or infrastructure) to see behind the polished veneer. The maps created are sometimes self-explanatory, or may require interpretation by someone with more familiarity with the domain, or actual meeting in which the map was created.

As an example, consider an extract from the Compendium map created during the ScholOnto project meeting to revise the ontology. A log of link type usage was imported into Compendium, which converted it to a visual tree structure. In Figure 3 below, the class of *taxonomic* links is being discussed, with annotations for each instance (e.g. *part of*), and associated notes and maps (on the right) containing more discussion on each link type.

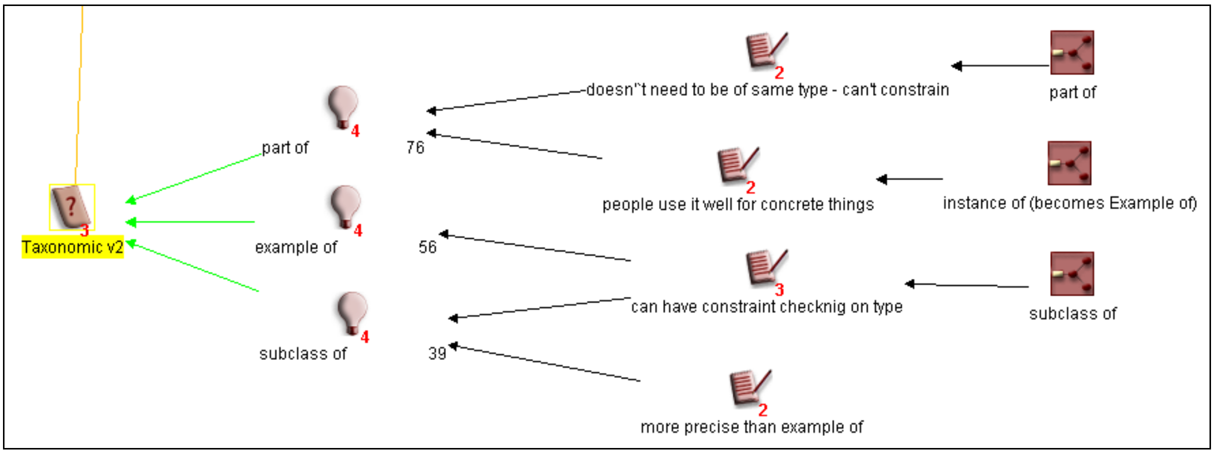

Figure 3: Use of one discourse mapping tool, Compendium, to support the collaborative redesign of ScholOnto's discourse ontology

In summary, what we are doing is applying one IDCP tool for real time discourse capture and structuring (Compendium) to support the design of another IDCP tool (ScholOnto). Whilst the example above is from Compendium's deployment for co-present meetings, it has also been used extensively to augment online meetings, providing a collaborative display to track discussion, summarise issues and export out to other tools [11].

8 Conclusion

We have taken a number of the IDCP themes emerging from the Paris workshop, and illustrated how they manifest in ScholOnto. This strengthens the case for these as conceptual tools for talking about IDCPs, and perhaps for comparing and contrasting examples.

In terms of broader research into IDCPs, we tentatively suggest the following generalisations from what we are learning in ScholOnto:

- **Spaces for interpretative work.** ScholOnto aims to provide a way to represent how a distributed research community is collectively interpreting (and contesting) the ideas proposed in its literature — what we call *claims*. But expressed more generically, the focus is on providing an explicit layer above a set of resources in an IDCP to engage in interpretative, sensemaking work on the meaning of those resources, with an explicit set of requirements to satisfy the heterogeneous nature of research discourse. This idea could generalise to other systems where work involves the explicit representation and debating of named objects and their relationships. This hypothesis will be tested as a range of user communities experiment with ScholOnto.

- **Claim structure.** The structure of a Claim for connecting concepts has withstood a first iteration of implementation, testing and revision. The decoupling of abstract relational classes from their rendering as natural language dialects has proven to have useful properties for both use and maintenance.

- **Structural analysis in IDCPs.** Although not an explicit principle in its initial conception, claim structures can be treated as graphs for certain purposes. An issue for discussion is whether graph theory and other structural analyses have other applications in IDCP research (we have already noted Social Network Analysis), and might serve as a way to integrate perspectives in a formal way.

9 Acknowledgements

The Scholarly Ontologies Project (2001-04) is funded by the UK's Engineering and Physical Sciences Research Council, under the Distributed Information Management Programme (GR/N35885/01).

10 References

1. Buckingham Shum, S., Motta, E. and Domingue, J. ScholOnto: An Ontology-Based Digital Library Server for Research Documents and Discourse. International Journal on Digital Libraries, 3, 3, 2000, pp. 237-248 [www.kmi.open.ac.uk/projects/scholonto].

2. Buckingham Shum, S. and Selvin, A.M. Structuring Discourse for Collective Interpretation. 1st International Workshop on Infrastructures for Distributed Collective Practices, Paris, 19-20 Sept., 2000 [www.limsi.fr/WkG/PCD2000].

3. Garfield, E. Research Fronts. Current Contents, October 10, 1994 [www.isinet.com/isi/hot/essays/citationanalysis/11.html].
4. Knott, A. and Mellish, C. A Feature-Based Account of Relations Signalled by Sentence and Clause Connectives. Language and Speech, 39, 2-3, 1996, pp. 143-183
5. Knott, A. and Sanders, T. The Classification of Coherence Relations and their Linguistic Markers: An Exploration of Two Languages. Journal of Pragmatics, 30, 1998, pp. 135-175
6. Egghe, L, and Rousseau, R. Co-Citation, Bibliographic Coupling and a Characterization of Lattice Citation Networks. Scientometrics (to appear),
7. Mancini, C. and Buckingham Shum, S. Cognitive Coherence Relations and Hypertext: From Cinematic Patterns to Scholarly Discourse. In Proc. ACM Hypertext 2001, Aug. 14-18, Århrus, Denmark, 2001, ACM Press: New York [www.kmi.open.ac.uk/tr/papers/kmi-tr-110.pdf].
8. Motta, E. Reusable Components for Knowledge Modelling. IOS Press: Amsterdam, NL, 1999
9. Erdos, P. and Renyi, A. On the Evolution of Random Graphs. Publications of the Mathematical Institute of the Hungarian Academy of Sciences, 5, 1960, pp. 17-61
10. Robinson, M. and Bannon, L. Questioning Representations. In: Proc. of ECSCW'91: 2nd European Conference on Computer-Supported Collaborative Work, Bannon, L., Robinson, M. and Schmidt, K., (Ed.), Kluwer: Amsterdam Sept 25-27, 1991, pp. 219-233
11. Selvin, A.M. and Buckingham Shum, S.J. Rapid Knowledge Construction: A Case Study in Corporate Contingency Planning Using Collaborative Hypermedia. Journal of Knowledge and Process Management, in press [http://kmi.open.ac.uk/tr/papers/kmi-tr-92.pdf].
12. Toulmin, S. The Uses of Argument. Cambridge University Press: Cambridge, 1958
13. Wasserman, S. and Faust, K. Social Network Analysis: Methods and Applications. Cambridge University Press: Cambridge, 1994
14. Watts, D.J. Small worlds : the dynamics of networks between order and randomness. Princeton University Press: Princeton, NJ, 1999