KMi

Knowledge Media Institute

Scholarly Discourse as Computable Structure

Simon Buckingham Shum, John Domingue and Enrico Motta

KMI-TR-93

25th April, 2000

Second International Workshop on Structural Computing
San Antonio, Texas, June 3, 2000 ***ACM Hypertext 2000*** **[www.ht00.org]**

The Open University

Second International Workshop on Structural Computing, San Antonio, Texas, June 3, 2000. *ACM Hypertext 2000* [www.ht00.org]

Scholarly Discourse as Computable Structure

Simon Buckingham Shum, John Domingue and Enrico Motta

Knowledge Media Institute, The Open University, Milton Keynes, MK7 6AA, U.K.
{S.Buckingham.Shum, J.B.Domingue, E.Motta}@open.ac.uk
http://kmi.open.ac.uk/projects/scholonto

Abstract. In their initial proposal for structural computing (SC), Nürnberg et al. [18] point to hypertext argumentation systems as an example of an application domain in which structure is of first-order importance. In this paper we summarise the goals and implementation of a knowledge based hypertext environment called ScholOnto (for Scholarly Ontologies), which aims to provide researchers with computational support in representing and analysing the structure of scholarly claims, argumentation and perspectives. A specialised web server will provide a medium for researchers to contest the significance of concepts and emergent structures. In so doing, participants construct an evolving structure that reflects a community's understandings of its field, and which can support computational services for scholars. Using structural analyses of scholarly argumentation, we consider the connections with structural computing, and propose a number of requirements for generic SC environments.

Introduction

Structural Computing (SC) has been proposed as a new paradigm which generalises hypertext's interest in explicit, computable structure to a philosophy for computing more widely. The approach is distinguished by its assertion of the primacy of structure over data: "Structure should be the ubiquitous, atomic building block available to all systems at all times and from which other abstractions (including data) are derived" ([18], p.96). Services for detecting, analysing and manipulating structure should therefore be represented using abstractions that transfer across domains and systems. We find SC an interesting proposal with strong connections to our own work. In this short paper, we outline a knowledge based digital library server currently in development, which focuses on the representation of scholarly claims and discourse as semantic structures (see [7] for details). Specifically, we show the connections between this specific domain and SC, and propose several requirements that analyses of our system motivate for generic SC systems.

Scholarly publishing, literature analysis, and ScholOnto

Representing discourse structure to assist scholarly analysis

We are concerned with future infrastructures for *scholarly publishing,* by which we include scientific, technical and medical research in both academia and industry. Scholarly publishing technologies are currently focused on using networks to access *digital analogues of paper, bibliographic metadata, and databases.* Whilst interoperable repositories make such information increasingly accessible, they provide no analytical leverage for *interpreting* the information. The power of the network (i) as a medium for scholarly discourse, and (ii) as a representation for conceptual structures and perspectives within a research community's digital library remains unexplored.

Support is needed for researchers who, typically, are interested in the following kinds of phenomena (exemplified with queries):

- The *intellectual lineage* of ideas: e.g. where has this come from, and has it already been done? ("Are there any arguments against the framework on which this paper builds?")
- The *impact* of ideas: e.g. what reaction was there to this, and has anyone built on it? ("Has anyone generalised method M to another domain?" "Has anyone extended Language L?")
- *Perspectives*: are there distinctive schools of thought on this issue? ("Has anyone proposed a similar solution to Problem P but from a different theoretical perspective?")
- *Inconsistencies*: e.g. is an approach consistent with its espoused theoretical foundations?; is there contradictory evidence to a claim? ("Are there groups building on Theory T, but who contradict each other?")
- *Convergences*: are different streams of research mutually reinforcing in interesting ways? ("Who else uses Data X in their arguments?")

Currently, researchers have no way to articulate such questions in a library, analogue or digital. Current metadata initiatives are focused on the encoding of *primary content attributes* to improve retrieval and interoperability. Inconsistencies and interpretations in encoding at this level are considered undesirable in order to assist machine processing. In contrast, these are precisely the features that a system needs to support the *interpreted, knowledge level,* as researchers contest the significance of data, and the concepts which it underpins. The ScholOnto project seeks to address the fundamental requirement for an ontology capable of supporting scholarly research communities in interpreting and discussing evolving ideas: overlaying *interpretations* of content, and supporting the emergence of (possibly conflicting) *perspectives*.

Hypertext argumentation

In their initial proposal for structural computing, Nürnberg et al. point to hypertext argumentation systems as an example of an applications domain in which structure is of first-order importance. Argumentation schemes make use of semantic networks of typed nodes such as *Claims, Arguments, Evidence* and *Theories* (e.g. [14], [15], [20]), and computer-supported collaborative argumentation (CSCA) continues as an active research and development field [11]. CSCA does however have important cognitive and social dimensions which if ignored, lead to end-user rejection of systems. Our own research into hypertext support for argumentation, e.g. [3-6] emphasises that any discourse structuring scheme intended for untrained users must carefully balance simplicity with expressive power, and provide computational services in order to balance the cost/benefit tradeoff. Applied to scholarly discourse, the social fabric of a research field will influence, and in turn be influenced by, the explicit declaration of claims and relationships between researchers' work. The vocabulary provided must be sensitive to this, and be customisable to the language of different communities.

Towards ontologies for scholarly discourse

Research disciplines are in constant flux and by definition lack consensus on many issues. Whilst this renders futile the idea of a 'master ontology/taxonomy' for a discipline, there does appear to be one stable dimension, namely scholarly discourse—the *way* in which new work is expressed and contested. Thus, it is hard to envisage when researchers will no longer need to make claims about a document's *contributions* (e.g. "this is a new theory, model, notation, software, evidence…"), or contest its *relationships* to other documents and ideas (e.g. "it applies, extends, predicts, refutes…").

Our approach provides an environment for scholars to make claims about concepts in documents, both their own and those of others. Decoupling concepts from claims about them is critical to supporting the emergence and co-existence of multiple *perspectives*. The kind of ontology that we are moving towards is outlined in Figure 1, suggesting concepts and relationships suitable for a wide range of disciplines. This generic scheme already enables inter-operability between different domains, but we also envisage that different disciplines might re-express or specialize concept or relational types, e.g. an experimental field might specialise *predicts* into *hypothesises*. Details and examples of modelling are presented in [7]. It is important to emphasise that our ontology is not merely a taxonomy or metadata scheme. We use a rich modelling language (OCML, see below) to provide both declarative and operational semantics for the concepts and relationships required to deliver ScholOnto's services (discussed below).

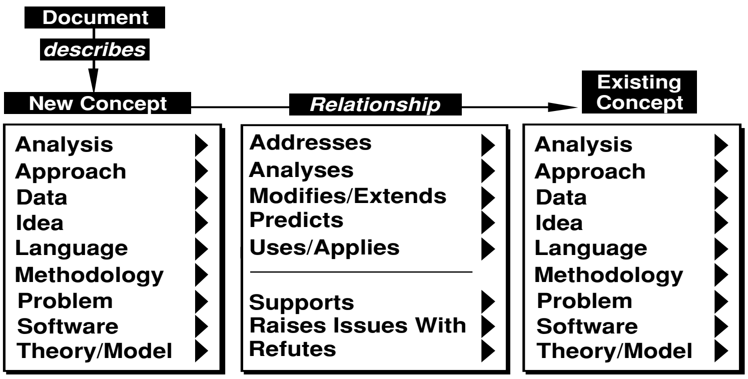

Fig. 1. Example ontology for making claims about the key contributions of a document, and its relationships to other concepts in the literature. It is hypothesised that such a scheme could be customised and adopted by research communities, whilst maintaining technical interoperability, and hence opening new possibilities for interdisciplinary discourse.

As a simple example, using a web form, a researcher *R1* might summarise a document's contributions as follows:

R1{{ *Language L* uses/applies *Model M*}{*Model M* modifies/extends *Theory T*}}

However, a researcher *R2* (or, significantly, a software agent monitoring the network) might challenge *R1*, in this case reasoning by a model of consistency grounded in the ontology's relational types:

R2{{*Model M* modifies/extends *Model N*}{*Model N* refutes *Theory T*} is-inconsistent-with {*Model M* modifies/extends *Theory T*}}

Structured discourse only pays off when the benefits outweigh the effort. In principle, discourse primitives such as those above would on their own support the emergence of useful networks and support a range of novel queries, but additional representational power is derived when they combine to form 'compound structures' corresponding to a range of interesting phenomena in a literature. Such structures could be declared meaningful by researchers, but in principle could also be identified by agents through structural computation.

Current status

ScholOnto is implemented as an application of several generic knowledge modelling technologies. Motta's *OCML* knowledge modelling language integrates both specification and operationalization, thus allowing seamless ontology specification, instantiation and integration with reasoning components [16], whilst Domingue's suite of Java tools [12], [17] provide graphical user interfaces onto an OCML knowledge base: *WebOnto* makes it possible to browse, edit, search, share and visualize knowledge models over the Web; *Knote* provides interactive forms to guide the process of populating the knowledge base, and *Lois* generates contextualised search forms for querying knowledge bases. A skeletal prototype has been implemented, and seeding with concepts has begun to validate the design of the ontology and to elicit requirements for a user interface for non-knowledge engineers.

Structural computing in ScholOnto

Knowledge based hypertext and structural computing

Implicit in ScholOnto (and SC) is the machine's ability to reason about structure. ScholOnto's knowledge base is a large, semantic network of typed concepts and relationships, plus rules for reasoning about the structure. The reasons for implementing in OCML (apart from local expertise) were the ease of defining semantic schema (as in Aquanet [15], or Germ [2], the meta-environment of which gIBIS was one application). It is thus simple to represent structural patterns of interest, and write inference rules (OCML is Lisp-based) to link system action to structure, or to compute new structures. Clearly there are many ways to implement structural reasoning capability, but we suggest that the overlap between knowledge based systems, semantic hypertext, and SC adds an interesting dimension to the SC initiative, especially where 'knowledge level' processing and interactive systems are combined.

The following sections illustrate the kinds of structural computing that ScholOnto will perform. With each example, broader implications for SC systems are highlighted as an 'SC requirement'. Whether these should be regarded as fundamental or merely desirable requirements is open for discussion.

Detecting contrasting perspectives

Perspectives within a literature could correspond to clusters of documents aligning for/against clusters of concepts. A ScholOnto agent could therefore be set to monitor the network for *contrasting perspectives* defined as structures where *≥3 documents support ideas of type A and challenge ideas of type B, and ≥3 documents do the converse*. The ontology thus provides the basis for detecting the emergence of 'camps' within a literature in which the basis for one cluster of documents is attacked by another cluster. If SC environments are to be customisable by non-programmers, structure and behaviours need to be accessible. One requirement might therefore be:

- **SC Requirement: High level scripting languages and user interfaces for end-users to define structures of interest (e.g. for searches or agents)**

The detection of structural corrollaries of perspectives would make use of *link families*. An example would be 'difference of opinion' with a concept or perspective, that is, relationships varying in strength from *raises issues with* or *is inconsistent with,* to *challenges* and *refutes.* Such a family could be usefully defined in order to cover relational sets that are of interest (e.g. "show me any papers that 'have problems' with this theory"). In addition, if a query is posed to the system that implicates the milder *raises issues with* relationship, it is reasonable for the system also to infer that the stronger *refutes* relationship will also be of interest, and use this to retrieve potentially relevant material. This level of knowledge also enables the computation of links that may not have been explicitly declared.

- **SC Requirement: Generalization and specialization of relational types (e.g. enabling reference via a parent to a whole set of relational types, or 'intelligent inference' based on known relationships between types)**

Link families are a specialised kind of composite structure that aggregate links.

Perspectives, and composite structures

Scholarly perspectives themselves emerge from, and come to represent, a configuration of concepts (that normally has established a following). An example might be the debate between those who subscribe broadly to symbolic processing approaches to AI, and those who work from a situated cognition perspective (for instance, we could imagine a properly elaborated model of the exchanges in [9]). Closer to home, the *As We Should Have Thought* paper seeks to establish *structural computing* as a coherent perspective on the design of systems by drawing on existing work, and the purpose of these SC workshops is to map out in more detail the conceptual roots of this perspective, and its ability to solve persistent problems. There would be a need within ScholOnto to be able to refer to *structural computing* as a perspective in a simple way that encapsulates this community. We are talking of course about the need for *composite nodes*, as proposed by Conklin [10] and reiterated recently by Anderson [1] in his SC analysis of software engineering environments (although composite nodes are more usefully conceived as composite structures within an SC paradigm).

- **SC Requirement: Composite structures that encapsulate sub-structure for ease of human/machine reference/processing.**

Arbitrary granularity: micro- and macro-argumentation

Argumentation/discourse is conducted at many levels of detail, so (in concert with composite structures) ScholOnto must support arbitrary granularities of structure (another of the SC requirements noted in [1]). At present, ScholOnto has been designed for 'macro-level' discourse (emphasising key claims and relationships *between* documents), bearing in mind the lessons from previous hypertext argumentation systems which indicate a reluctance or inability on the part of untrained users to make the structure of too many ideas and concepts explicit. Even at this level, we envisage the need for shades of discourse level (implied by the need for composites). However, closely related work on hypertextual argumentation by Kolb [13] and Carter [8] describes the reification of argumentation structure at a finer granularity, suggesting possible uses of ScholOnto if there was the demand for its ontology to be refined to support fine-grained argumentation. This corresponds to a requirement to represent *intra-document* as well as inter-document structure. To summarise,

- **SC Requirement: Support the expression of multiple levels of structural granularity, enabling common structural reasoning capabilities to be applied at multiple levels of analysis.**

Conflicting structures

A research discipline's raison d'être is to debate and evaluate conflicting views. As a medium for making claims and engaging in argumentation, ScholOnto must support this process, and thus is fundamentally perspectival in philosophy. The ontology therefore permits conflicting links to be made between concepts, regarding them as *claims* associated with a person, and open to contest. This is of course very different to systems seeking to maintain structures for machine interpretation only, or for applications where inconsistency is undesirable. Current web metadata and ontological annotation initiatives fall into this category, being focused on the encoding of primary content attributes to improve retrieval and interoperability. Inconsistencies and interpretations in encoding at this level are considered undesirable in order to assist machine processing. In contrast, these are precisely the features that ScholOnto needs for human interpretation.

What are the implications for generic SC environments? There are no doubt other application domains where multiple, logically conflicting, structures need to co-exist. A collaborative SC environment should support this, at least when end-users expect plurality of perspectives, and need to negotiate the meaning of structures. Avoiding premature commitment to structure was, after all, precisely the motivation behind spatial hypertext systems such as VIKI [19] which are a prime application for structural computing.

- **SC Requirement: Support the emergence, co-existence, and analysis of, structures contributed by multiple authors, which may conflict logically.**

Structural patterns signalling inconsistency

With many research groups working with shared concepts, possibly across different disciplines, these concepts are invariably misunderstood, or mutate as they are better understood. Authors may ignore, or not be aware of each other (publications might be years apart, or in different fields), but ScholOnto's model of the network enables one to trace consistency (as defined by the ontology). It would be useful for researchers to explore 'what if' scenarios for new ideas (for instance publishing a draft concept map in a personal space to see what the system came back with in terms of inconsistencies or related work). Or, we can imagine a journal reviewer testing the ideas in a paper for consistency with the concepts on which it claims to build.

The ScholOnto functionality deployed here is the ability to use the ontology to define arbitrary principles (e.g. "consistency"), and heuristics to detect possible violations of that principle (e.g. "existence of both positive and negative relationships between two concepts"—where link families have been defined, as discussed above). The generic SC requirement motivated by this example might be expressed as:

- **SC Requirement: Support the definition of arbitrary criteria and provide a means to validate structures against those criteria.**

Conclusion

To summarise, our analysis of the computational work that the ScholOnto system must perform confirms that it is an application that falls squarely in the domain of structural computing. Our representation, through the use of an ontology for scholarly claims and argumentation, highlights the fruitful overlap between knowledge based hypertext and structural computing, and raises for discussion several new requirements for generic structural computing environments.

References

1. Anderson, K. Software Engineering Requirements for Structural Computing. In *First International Workshop on Structural Computing*, Darmstadt (Feb. 21), 1999, Technical Report CS-99-xx, Dept. Computer Science, Aarhus University, DK <http://www.daimi.au.dk/~pnuern/sc1/submissions/anderson.html>

2. Bruns, G. Germ: A Metasystem for Browsing and Editing. Microelectronics and Computer Technology Corporation, *Technical Report STP-122-88*, 1988

3. Buckingham Shum, S. Analyzing the Usability of a Design Rationale Notation. In *Design Rationale: Concepts, Techniques, and Use,* Moran, T.P. and Carroll, J.M., (Ed.), Lawrence Erlbaum Associates: Hillsdale, NJ, 1996, pp. 185-215

4. Buckingham Shum, S. Negotiating the Construction and Reconstruction of Organisational Memories. *Journal of Universal Computer Science (Special Issue on Information Technology for Knowledge Management),* 3, 8, 1997, pp. 899-928 <http://www.iicm.edu/jucs_3_8/>

5. Buckingham Shum, S. and Hammond, N. Argumentation-Based Design Rationale: What Use at What Cost? *International Journal of Human-Computer Studies*, 40, 4, 1994, pp. 603-652

6. Buckingham Shum, S., MacLean, A., Bellotti, V. and Hammond, N. Graphical Argumentation and Design Cognition. *Human-Computer Interaction*, 12, 3, 1997, pp. 267-300 <http://kmi.open.ac.uk/kmi-abstracts/kmi-tr-25-abstract.html>

7. Buckingham Shum, S., Motta, E. and Domingue, J. Representing Scholarly Claims in Internet Digital Libraries: A Knowledge Modelling Approach. In *Proc. of ECDL'99: Third European Conference on Research and Advanced Technology for Digital Libraries*, Paris, Sept. 22-24, 1999, Springer-Verlag (LNCS 1696), pp. 423-442 <http://kmi.open.ac.uk/projects/scholonto/>

8. Carter, L.M. Arguments in Hypertext. In *Proc. Hypertext 2000*, San Antonio, TX, 2000, ACM: New York

9. Cognitive Science Journal Special Issue on Symbolic Reasoning and Situated Action. *Cognitive Science*, 17, 1993, pp. 1-133

10. Conklin, J. and Begeman, M.L. gIBIS: A Tool for All Reasons. *Journal of the American Society for Information Science*, 40, 1989, pp. 200-213

11. CSCA: *Computer-Supported Collaborative Argumentation Resource Site*. Knowledge Media Institute, Open University, UK <http://kmi.open.ac.uk/sbs/csca>

12. Domingue, J. and Motta, E. PlanetOnto: From News Publishing to Integrated Knowledge Management Support. *IEEE Intelligent Systems (Special Issue on Knowledge Management and Knowledge Distribution over the Internet)*, , in press

13. Kolb, D. Scholarly Hypertext: Self-Represented Complexity. In *Proceedings of The Eighth ACM Conference on Hypertext*, Southampton, 1997, pp. 29-37

14. Lee, J. SIBYL: A Tool for Managing Group Design Rationale. In *Computer Supported Cooperative Work*, Los Angeles, CA, 1990, ACM Press: New York, pp. 79-92

15. Marshall, C.C. and Rogers, R.A. Two Years before the Mist: Experiences with Aquanet. In *Proceedings of the Fourth ACM Conference on Hypertext*, 1992, pp. 53-62

16. Motta, E. *Reusable Components for Knowledge Modelling*. IOS Press: Amsterdam, NL, 1999

17. Motta, E., Buckingham Shum, S. and Domingue, J. Ontology-Driven Document Enrichment: Principles and Case Studies. *International Journal of Human-Computer Studies*, , 2000, in press <http://kmi.open.ac.uk/projects/scholonto/>

18. Nürnberg, P.J., Leggett, J.J. and Schneider, E.R. As We Should Have Thought. In *Proceedings of Hypertext'97: 8th ACM Conference on Hypertext*, Southampton, 1997, pp. 96-101 <http://journals.ecs.soton.ac.uk/~lac/ht97/pdfs/nuern.pdf>

19. Shipman, F.M. and Marshall, C.C. Formality Considered Harmful: Experiences, Emerging Themes, and Directions on the Use of Formal Representations in Interactive Systems. *Computer Supported Cooperative Work*, 8, 4, 1999, pp. 333-352

20. Trigg, R. and Weiser, M. TEXTNET: A Network-Based Approach to Text Handling. *ACM Transactions on Office Information Systems*, 4, 1, 1983