

Formalization, User Strategy and Interaction Design: Users' Behaviour with Discourse Tagging Semantics

Bertrand Sereno,¹ Simon Buckingham Shum² and Enrico Motta²

Centre for Advanced Learning Technologies, INSEAD
Boulevard de Constance, F-77305 Fontainebleau
France
bertrand.sereno@insead.edu

Knowledge Media Institute, The Open University
Walton Hall, Milton Keynes
UK
sbs@acm.org / e.motta@open.ac.uk

ABSTRACT

When authors publish their interpretations of the ideas, opinions, claims or rebuttals in the literature, they are drawing on a repertoire of well understood moves, contributing to an extended discourse. Readers also bring their own perspective to documents, interpreting them in the light of their own research interests, and initiating, for instance, new connections that may not have been intended by authors. Collaborative, social, tagging holds promise as an approach to mediating these processes via the Web, but may lack the discourse dimension that is fundamental to the articulation of interpretations. We therefore take a hybrid semiformal approach to add structure to freeform folksonomies.

Our experience demonstrates that this particular brand of tagging requires tools designed specifically for this *sensemaking* task by providing enough support to initiate the annotation, while not overwhelming users with suggestions. We describe a tool called ClaimSpotter that aims at supporting this tradeoff, through a novel combination of system-initiated tag recommendations, Web interface design, and an expanded conception of how tags can be both expressed, and semantically linked. We then report a detailed study which analysed the tool's usability and the tag structures created, contributing to our understanding of the implications of adding structure to collaborative tagging.

Categories and Subject Descriptors

H.1.2 [Information Systems]: User/Machine Systems—*Human information processing*; H.4.m [Information Systems]: Miscellaneous

Keywords

Social tagging, Sensemaking, Discourse Relations, Semantics, Argumentation, Usability, Pragmatic Web

1 INTRODUCTION

Our communities, local, national and international, are confronted by problems that are complex due to the changing environment, incomplete or ambiguous information, and stakeholders with different perspectives. Such domains include strategic planning in business, government policy formulation, time-pressured mission operations, and almost all scholarly research. The *sensemaking* activity that these contexts demand [25], requires analysts to construct plausible narratives that frame the problem, account for the available evidence, and motivate action, enabling “an openly reflexive forum in which communities of knowing explicitly talk about their understandings” [2]. Progress is made by making moves that express and contest interpretations of the world, although these different contexts clearly have very different genres of discourse and criteria for acceptance. The focus of this paper is on the work of academic researchers, but we argue that related work shows that an approach grounded in discourse relations is applicable to a broader range of applications.

It is established from corpus analyses that when researchers publish their interpretations of the ideas, opinions, claims or rebuttals in the literature, they are drawing on a repertoire of well understood moves, contributing to an extended discourse [21]. Although the Internet is accelerating the pace of exchanges, scholarly and scientific discourse still proceeds in the shadow of the printing press, with exchanges now disseminated as *digital* prose. While information retrieval and text analysis technologies help to infer certain kinds of structure within and between papers, our research is complementary, exploring a ‘network-native’ paradigm in which the key claims made by an author (and the interpretations made by their readers) are published as explicit new connections to the literature. The research question driving our research is: *can we model the discourse structures we find in research communities as explicit structures, and if so, what support*

Work conducted while the first author was at The Open University.
Copyright is held by the author/owner(s).
WWW2007, May 8-12, 2007, Banff, Canada.

tools can we provide to construct, navigate, and interrogate such structures? Such approaches to knowledge publishing and negotiation on the Web will ideally be both quickly learnable, yet sufficiently expressive to permit researchers to make important scholarly moves, and assist them in making sense of the emergent structures at scale.

What is especially interesting about scholarly discourse is the fact that “the truth” or “the significance” of any claim is open to contest. While this may be extreme in the case of philosophy and the humanities, it is self-evidently also the case even in computing and the hard sciences. There is no single reading of a paper; interpretations may differ significantly between readers and authors (hence the need for peer review); readers bring their own unique perspective to a paper, seeing new connections that the author may never have intended. Seeing old things in new ways is the essence of creativity. This is the orientation we bring to harnessing the power of social tagging, but augmented with discourse semantics, as we strive to create effective infrastructure for scholars to express—and contest—claims to knowledge.

2 SCHOLARLY TAGGING

2.1 Questions no search engine can answer

Consider the following questions that interest students and researchers, but which neither Internet search engines nor domain-specific digital libraries can assist in answering: *What data refutes this hypothesis? Are there different schools of thought in this field? Is there an analogy between this process in fields X and Y? Why does this paper cite that one? How did these contrasting perspectives interpret this result?* The answers to these questions are grounded in the discourse moves that researchers make in their writing: the arguments, rhetoric and positioning of their claims with respect to the literature. In our present infrastructures, these are questions that can only be answered by reading the paper, although there is active research on the automated analysis of argumentative relationships between papers [22].

These are fundamentally issues of interpretation, which fall outside ontology-based Semantic Web approaches which model stable, consensus, ‘objective’ worlds (albeit always from a perspective). Nor can they be answered by scientometrics (e.g. citation analysis) which do not have enough insight into the nature of the moves being made. We are now squarely in the realm of pragmatics, where meaning derives from interpretation, perspective, contextualisation and argumentation—in other words, the construction of plausible narrative, as introduced at the start.

2.2 Discourse semantics for annotating claims

We take a hybrid, semiformal approach to add structure to freeform folksonomies. Details can be found in [4, 13, 24]¹

¹ The work of ISO/TC37/SC4 shares a common interest in discourse and coherence relations: <http://www.tc37sc4.org>

1. As with folksonomies, tags remain unconstrained freetext strings, although users can choose to take care to reuse existing tags in order to increase the visibility of their tagging, or to discover new connections. In our context, however, tags may become phrases or even a sentence or two if they are used to express, for instance, a hypothesis, a prediction or a research result.
2. A critical difference is that tags may be linked not just to a URI, but to each other. We term a tag—*relationship*—tag triple a *claim*, that is, a meaningful connection being asserted between two ideas. A claim may also link from/to other claims, as the ideas grow in complexity. A claim is also directed: it has a source and a destination tag.
3. Tags are linked using a typology derived from argumentation and the most common moves made in research publications. Users select the relationship from a menu of predefined relationships (e.g. *is consistent with*, *refutes*, *addresses*, *solves*, *improves on*, *is analogous to*, *uses/applies*).
4. Tags may optionally be classified (e.g. *problem*, *evidence*, *data*, *method*, *theory*), but these are pragmatic, contextual roles, holding only in the context of a particular claim. Thus, in one context, a research result might be a *problem*, while, in another context, it might be an *assumption*.

2.3 Relation to our previous work

Elsewhere [4, 24], we have demonstrated how a digital library can be tagged in this way with annotation tools, the resulting network navigated via interactive visualizations, and the semantic searches enabled by modelling discourse relations (e.g. show papers that *support*, or *contrast* in some way with this paper; show the *lineage* or *ancestors* of the idea represented by this tag). We have also evaluated how students make use of some of the tools to navigate and search prepopulated networks modelling a literature [4].²

Having demonstrated the potential of scholarly publishing and annotation using discourse relations to annotate texts, the challenge (as for any structured knowledge capture tool) was *can users do this?* To date, we have not presented data on tag *authoring* behaviour. This paper reports the first quantitative and qualitative analysis of the ways that novices and experts approached semantic tagging in their first encounter with a software tool. Semantic tagging behaviour is inextricably linked to (1) the semantic scheme, as introduced above, and (2) the user interface and functionality of the tagging tool, introduced next.

2.4 ClaimSpotter

The previous annotation tools developed to support our tagging approach did not allow direct annotation of the target document. This was an explicit goal with

² Demonstrations and screencasts: <http://claimmaker.open.ac.uk>

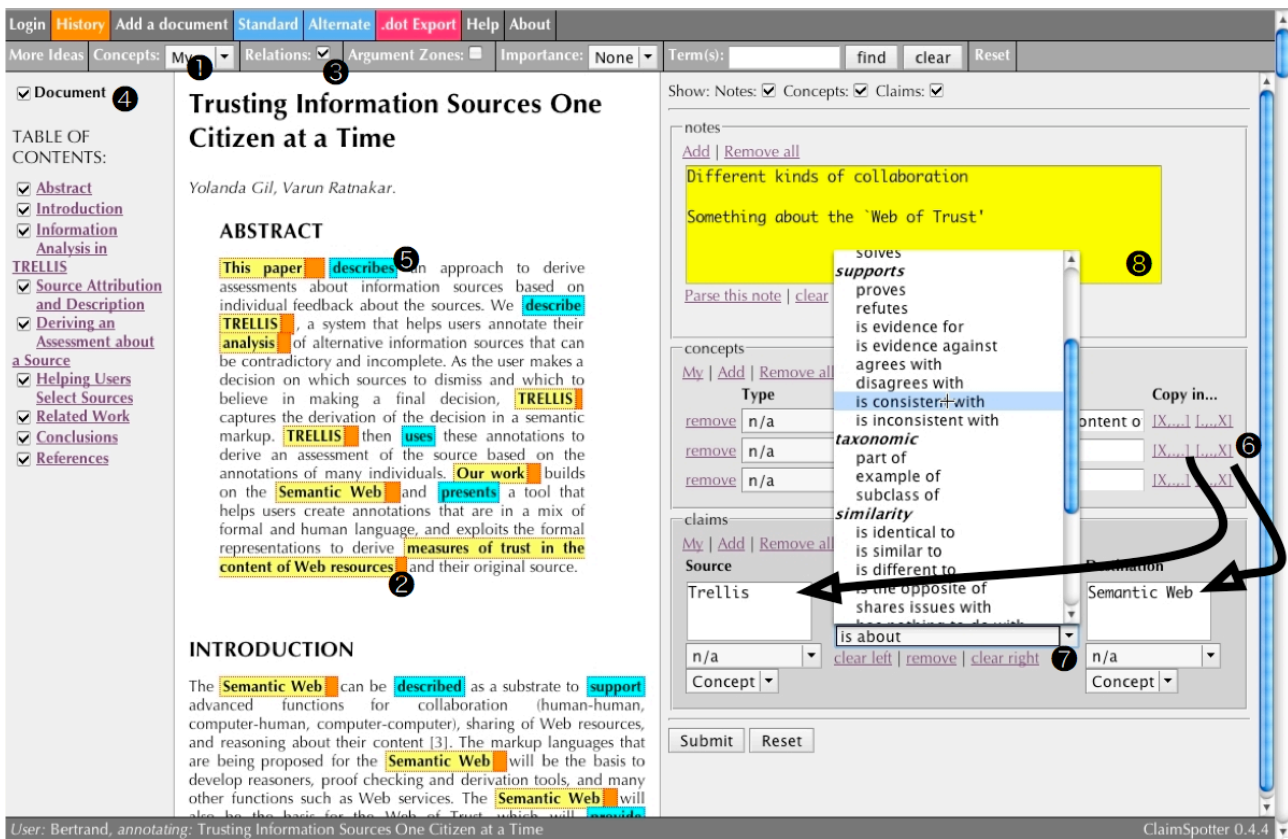


Figure 1: The ClaimSpotter interface. Key: [1] The My/All Tags toolbar button highlights text matching tags on this paper (just the user's, or all tags), e.g. [2] *measures of trust in the content of Web resources*. Clicking a highlighted tag enters it in the tag linking form. Similarly, the Relations filter [3] optionally applied to the whole document [4], highlights verbs matching, or synonymous with, the link types, e.g. [5] the verb *describes* in the text has been matched to the tag relation *is about*. Tag triples can be built from existing tags using the [X,..] and [,..,X] buttons [6] to specify the left and right sides of the triple. The tag link is selected from the menu [7]. Notes can be saved as tags [8], which like the document text, can on request be parsed for matching tags and relational types.

Further examples and screencasts: <http://kmi.open.ac.uk/projects/hyperdiscourse/tools/claimspotter>

ClaimSpotter, designed to support document sensemaking tasks: reading, highlighting areas of potential interest, making notes, looking for specific kinds of papers in the bibliography, and so forth. While researchers clearly do this all the time on paper, or with freetext annotations in various document viewers, the challenge was to support users in these tasks with our semantic tagging approach.

ClaimSpotter's design aims to initiate and sustain a dialogue between annotators and the target document, via (i) content-based support for tagging, in the form of *recommendations*, and (ii) an interface displaying these recommendations overlaid on the text (cf. figure 1). Details are in [18, 19]; we turn now to the evaluation study.

3 USER EVALUATION STUDY

There are many types of evaluation. A summative analysis could evaluate technical performance (e.g. of

recommendation agents), or characterise the impact of a tool on the practices of researchers (which one might do with mature tools like Google, Wikipedia or del.icio.us). We conducted a formative evaluation of a new prototype, in order to develop a language in which to describe as yet poorly understood phenomena. Our specific objective was to characterise how annotators approached the task we gave them with an unfamiliar tool, paying particular attention to how the affordances of the user interface (that is, the visual cues it provided for interaction) shaped tagging behaviour, summarised quantitatively against various measures, and explained through qualitative coding of the data.

3.1 Methodology

We recruited 13 annotators (referred to as a1–a13) who used ClaimSpotter to annotate a 2 page research paper which they had preferably authored, or were at least very familiar with to avoid any comprehension problems. Ten users were PhD

students, two were research fellows and the last was a professor. None had used ClaimSpotter before. Four of them (a1–a4: 1 student, 2 RFs and the professor) were considered ‘experts’ with the tag linking scheme, being members of the project team. The remaining nine (a5–a13) were considered ‘beginners.’

Each session was limited to one hour. Screen interactions were recorded with a capture tool, and all comments and discussions recorded, resulting in high quality audio-visual data as digital movies. A tutor (first author) bootstrapped each annotation process by defining a few tags for each document. He was also present throughout the session to provide assistance when needed, but also to engage discussion when suggestions were made. A questionnaire sent one week after the experiment was designed to elicit opinions on the main strengths and weaknesses of the interface, and on the ways it could be improved. See [18] for more details.

4 QUANTITATIVE ANALYSIS

257 tags and 160 claims were submitted, giving on average 19.8 tags and 12.3 claims per annotator, with no major difference between the 4 experts and 9 beginners, the former entering marginally more tags (a mean of 20.75 against 19.3) and links (a mean of 14.75 against 11.2) than the latter.

4.1 Tags

Most tags submitted were 1–3 words. 164 out of 257 tags (64%) submitted were ≤ 3 words. Short tags (representing proper nouns, acronyms or projects names) were as frequently submitted by novices as by experts. Most of these tags were used twice, while a handful were used three times. Duplicated tags were either created ‘explicitly’ by reusing a tag previously created in the current document, or ‘implicitly’ by typing a text string which happened to be already used as a tag. However, the documents chosen by participants were so different that duplicates were mostly due to annotators reusing a tag created beforehand by the tutor. We also noticed that reused tags were not necessarily composed of short tags only: some longer tags were reused.

4.2 Tag triples (“claims”)

22 relation types (out of the 36 available) were used. 7 out of these 22 were used only once or twice. ‘General’ relations were the most frequently used ones, but it is difficult to talk about these most frequently used relations, as the papers considered were different. A more interesting aspect may be to identify which relations were the most consistently used by annotators.

The relations *uses/applies/is enabled by* and *is about* were the two most consistently used: only 3 annotators did not use the former at all, and only 4 did not use the latter.

The examples below demonstrate the variety of tag triples created by participants:

[Domain ontology, ***is about***, A hierarchy of URIs on multiple levels]

[Universal physical access, ***is unlikely to affect***, Digital divide]

[Hypertext node juxtaposition. ***is analogous to***, Cinematic shot juxtaposition]

[(Evidence) In the Bristol trial, the awareness of the presence of other players was correlated with how much our participants enjoyed the game as well as with how engaged they felt, ***is consistent with***, Presence awareness of many other people is capable of causing, feel good factor]

[Magpie moves away from hypermedia towards open service-based architectures, ***is evidence for***, [Magpie, ***improves on***, COHSE]]

It can be seen that tags ranged from single words to a sentence, are optionally given a type (cf. fourth example.) In the last example, a tag is linked to another triple to create a compound claim.

4.3 The *is about* link

If we consider conventional tagging on the Web, the assignment of a tag to a URI is semantically very close to simply asserting that the content *is about* that tag. We performed a detailed evaluation of the use of the *is about* link, since it was one of the most commonly used links. It is what we might term a ‘less committing’ link compared to stronger, more argumentative relations such as *challenges*, *proves*, or *is analogous to*. This of course does not mean that *is about* links have little value: they have as much value as current tagging practices, and when used between two tags, such a connection can express a valuable and surprising stance if they were previously thought unrelated.

Experts submitted proportionally fewer *is about* links than beginners, which we attribute to their greater awareness of the other links available. Beginners, by contrast, were more likely to use *is about* as a placeholder ‘catch all’ link, especially when they had not yet established if the link they had in mind was on the menu (see the user strategy ‘Starting from the tags’ discussed shortly.)

Those annotators who made more links made proportionately more *is about* links. In contrast, annotators who made fewer links made almost no use at all of them. It appears that they focused directly on forging stronger links.

If we divide each annotator’s total link set in half, we find more *is about* links in the first half than in the second half. We interpret this as confirming the idea that this lower commitment link helped to scaffold users into this new mode of tagging. 8 annotators out of 13 had submitted at least one *is about* link. As they became more knowledgeable about the process and the links available, there seemed to be less need to fall back on *is about*. It can therefore be seen as a mechanism to incrementally formalize [20] one’s tagging. We can imagine ClaimSpotter prompting annotators at a later stage to review whether to ‘upgrade’ *is about* links to more specific ones.

5 QUALITATIVE ANALYSIS

The qualitative analysis focused on the audio-video data. We used a shallow Grounded Theory methodology to code the video transcripts (to create concepts) and organise them (in order to draw relations between these concepts) [7]. The outcome of this methodology was (in Grounded Theory terms) a ‘theory’, that is, a set of plausible relationships holding among multiple concepts. Concepts emerged from the analysis and were constantly compared against each other through specialization of codes into sub-codes, or vice-versa, consolidating sub-codes into parents (called categories). Finally, a stable state (the point of theoretical saturation) was reached where the codes were judged to account for the salient phenomena. The final taxonomy is given in Table 1, providing a more nuanced vocabulary than available prior to the study, in which to describe users’ tagging behaviour with ClaimSpotter. Discussion is organised around the three top level themes: *Formalization*, *User Strategy* and *Interaction Design*.

5.1 Theme 1: Formalization

The analysis of behaviours grouped under *formalization* yield insights into the degree of cognitive effort it took users to use the new structured tagging scheme.

Assigning types to tags

Most users decided not to add a tag type simply because it was optional: types were assigned 34 times, out of a total of 257 tags. Twice, types were explicitly not assigned because there were too many (“*The interesting thing is that this specific example (tag) could fall in different categories.*”) and once, because there were not enough (“*It’s not a problem, it’s not a solution, and it’s not a methodology. I’d like something that says research field*”). Search was not part of this evaluation task (the focus of a previous experiment [23]. We have not yet gathered longitudinal data with extensive tag authoring and searching, but we hypothesise that as users learn that they can search on types (e.g. find all instances where this tag was considered an *assumption*), they might start to assign them in anticipation. This is analogous to expert users formulating compound specializations of tags in Web social bookmarking. Users

Formalization

- creating a tag
 - choosing a tag type
 - appropriate tag type
 - not perfect tag type but problem with or lack of a tag type
 - cannot find a tag type
 - removes tag type
 - deletes tag
- creating a claim
 - choosing a relation
 - removing a claim

...

discussion about formalism

User Strategy

- keeping things simple
- reducing amount of information on screen
 - looking for ideas
 - focussing on a particular area
 - hiding an area

...

- starting a claim from the tags
- starting a claim from the relation
- typing or selecting a tag
- incremental formalization
 - reusing a tag or a claim previously submitted

...

Interaction Design

- consistency
- feedback

...

Miscellaneous

Table 1: Extract of the data coding scheme which emerged from the analysis of tagging behaviour: themes, categories, sub-categories and codes

will willingly add tag complexity as it serves their anticipated needs.

Relation types

An appropriate relation was found in 115 occurrences, out of 160 total. However, just as we found with choosing a *tag* type, we observed difficulties in choosing a relation type:

- On 8 occasions, a ‘good enough’ relation was found. This means that the annotator kept and submitted the triple, although it did not express completely what she had in mind (“*I can say is similar to, since there is nothing else better than that*”);
- On 6 occasions, the problem was even more acute: “*The relation (that I want) is not there. So what do we do?*” It resulted in the removal of the whole triple that was being created.

Multiple attempts were sometimes needed to get a claim right. This implied either trying different relations and finding out which one looked (and, actually, sounded, as annotators were saying them aloud very often) best, flipping

the source and destination tags, or reformulating a tag to make it suit a given relation. We recorded 11 incidents when an annotator had to reformulate the wording of a tag because of a relation.

'Good' and 'bad' tags

One annotator commented that a tag she was considering adding was “a silly tag” but that she would “make it anyway”, because it was of interest to her. She then added: “I’m not sure if that tag’s going to be good. Maybe some of these tags are less useful than the others.” Prompted to comment on her notion of tag utility, her answer was most interesting: “A good tag will be something that is consistent, something that would appear again and again in the document. [Tag name] is a good tag for instance, compared to something I would use only once.” This notion of quality derived from potential reusability, which is clearly the conventional understanding that users bring to tagging. This puts a premium on short tags referencing real world entities, such as the names of theories, algorithms, problems or methods. These are, of course, the sorts of entities that are extractable automatically, compared to the more complex tags that ClaimSpotter supports, but which were more novel to users and were used less frequently. In devising an interface for more subjective interpretative tags, this comment gave us pause for reflection on how the interface could have encouraged richer tags, to move users beyond the stereotype. See also the later discussion on the bias we unwittingly gave in the user interface to short, matched tags, which reinforced this emphasis.

5.2 Theme 2: User Strategy

Users are hard to predict, each brings his/her own unique knowledge of their domain, and varying expectations about the formalism and tool. Although we might have expected as many strategies as we had annotators, we believe we have identified several patterns.

Roles played by recommendations

We noted a difference in the amount of support annotators wanted from the interface, and its ability to extract and ‘recommend’ elements through text highlighting. Beginner annotator a7 made little use of the recommendations and spent most of the experiment inputting her own tags and claims, while all the other participants did actually use the suggestions.

Expert annotator a1 preferred at one point to deactivate the suggestions because, in her words, “I don’t want to be too distracted by having too many things going on. At the moment, it seems to be quite complicated. I’d rather keep it simple.” Later, however, she made use of the recommendations “to see if there’s anything inspirational (in this part of the document)”.

Recommendations were typically used to reduce the document to a set of potentially interesting focal fragments.

They could also be activated to discover (and reuse) existing tags, to position an argument with respect to peers’ tags, to find out how a particular tag was used over the corpus, to find peers’ tags and claims, to indicate which tags were associated to a cited document, or to indicate how a cited document was assessed by its author.

Incremental formalization

Tags and claims were not necessarily submitted immediately. Instead, they were often kept on the screen because annotators felt the need to see them to facilitate the creation of claims. Saying aloud the relations was also a phenomenon we often noticed, as mentioned earlier.

Another strategy-related phenomenon was related to the order in which annotators accessed the different resources at their disposal. They seemed to focus first on making their own annotations (possibly to get their feet wet with the formalism) before browsing through the history and looking for relevant tags and claims from their peers. This may have been an experimental artefact (the need to ‘get something done’ by the hour)—“For the time given, the easiest thing is to see the system suggestions and make your own. Because go back and look through the history may just take too much time”. But it may also have to do with a desire to appropriate the document first, to make it their own, before turning to what their peers said about it.

Starting from a relation vs. starting from tags

We also observed a striking difference between how (mostly) experts started from the relation type they wanted to use for a claim and how (mostly) beginners started from the two tags they wanted to put in relation, without knowing if the relation type they wanted to use existed. On reflection, this phenomenon is not surprising, but this was the first empirical evidence we had.

Towards a new kind of annotation process?

Although a few of the claim-spotting filters did exhibit some unwanted results on the papers provided by the annotators, the visual noise levels were not as damaging as had been feared. What was of interest to us was whether the very presence of recommendations and peers’ tags shaped annotators’ behaviour.

We characterise the effect that highlighted tags had as follows. From a situation in which annotators are given no cues as to how to tag a document, we moved to a situation in which they had to decide if an existing tag was good material to make a tag or a claim or not. We felt that annotation moved towards making a Yes/No decision in response to each recommendation. We will revisit this point later.

5.3 Theme 3: Interaction Design

We studied in detail the annotators’ interactions with the interface, and concluded that the environment was

reasonably intuitive within the constraints of the recorded task of annotating a single document. Longitudinal evaluation with large tag sets will undoubtedly reveal other design weaknesses.

Successful features

The presence of pull-down menus of tag types and relations on the screen succeeded as a visual scaffold: “*I’m looking through the types because I’m not familiar with them.*” The presence of the multiple tag types available also drew annotators’ attention to specific aspects of the paper that they might choose to focus on, e.g. what is the *problem* tackled, or the *methodology* proposed?

The tag-linking features were also very successful, encouraging a playful approach: the act of combining and swapping tags between the left and right sides of the link was made easier by not having to retype them, and introduced a *bricolage* aspect that encouraged experimentation.

As users gain confidence with a tool, they develop interaction routines, that is, compilations of micro-actions. These routines provide us with another way to describe the coupling between user interface and structured tagging.

Navigating and tagging by document section

A simple routine was navigating via the contents menu to a particular section, reading/skimming it and summarising it via a tag. This enabled a user to work through the text systematically, and confirmed the value of integrating the document and the annotation in a seamless interface.

Navigating and tagging by recommendation

A variation on this was to work from the output of filters: switching on a filter, looking at a highlighted area in the document, reflecting on it, modelling a tag or a claim, and moving to the next highlighted area. This sequence again confirmed the ability to move fluidly between engaging with the document, and tagging, with highlighted tags in the text acting as attention-catchers.

Combining tags into claims

The process of claim authoring evolved into a recognisable pattern of creating a tag, creating another tag, combining them in a claim, looking for a discourse link, not finding one, flipping the order of the tags in the relation, and finally finding an appropriate relation.

Reusing and adapting peers’ tags

Some users learnt to use the less obviously available history window (listing, among others, non-matched tags.) Consulting the tags available and reusing one or more in one’s own tag space demonstrated that annotators did benefit from peers’ tags.

Annotating and checking for visual feedback

This move illustrates the dominance of ‘visible’ tags, which as discussed, had not been foreseen. Users would select, copy and paste some text from the document into a tag, submit it, and immediately activate the ‘my tags’ filter to see it appear highlighted in the text, confirming that it had been recorded.

6 DESIGN WEAKNESSES

In this section, we reflect on some of ClaimSpotter’s design weaknesses, and consider improvements that may also be of relevance to other collaborative knowledge structuring tools.

6.1 Information overload?

ClaimSpotter’s filters were designed to address the challenge of supporting an annotator in the task of locating and tagging a document’s contributions. The presence of highlighted tags and text fragments undoubtedly shaped the annotation process, and we have evidence that annotators valued seeing these, with some variations in when they activated them. We did find evidence, however, that there may have been too much information. As mentioned by a4, “*the problem is, do you make your own claims, do you follow the system, do you go back to the history to see what the other people have said?*”

There is no question that for a one-hour experiment, there was indeed a lot of information to understand and digest. More studies are needed to introduce the different sources of support more gradually, and to let annotators decide which ones work best for them. Better ways to organise these recommendations need to be devised (work has begun on a dialogue assistant that helps annotators ask themselves focused questions about the document, and which suggests recommendations for each question).

6.2 ‘Current-document centeredness’

New users will focus on what they are offered by the display. ClaimSpotter’s document-centric design emphasised the current document, at the expense of easy access to cited documents, for instance. Our conclusion is that this resulted in a limited number of claims being made which connected tags originating in different documents. However, our other work has evaluated user interfaces that foreground the tag space structure, providing a complementary perspective [23].

6.3 User ‘laziness’

Our objective was to devise a more active interface to suggest possible tags. We now play devil’s advocate and ask if tags and claims would not be more reflective if they had to be devised manually by the annotator? By saving the annotator the cognitive effort of formulating their own tags, are we undermining the very process we want to promote?

We observed a tendency to create (i.e. reuse) tags from text fragments highlighted in the document by the recommendation filters. Some of these were copied and edited to taste, but they were nevertheless heavily inspired by the highlighted elements in the original document. While this seems to be a ‘good’ thing both in terms of usability (it lowers the barrier for constructing semantic literature models), and in terms of the building of a network promoting the reuse of tags, there is the corresponding risk that less effort is put into the annotation: the user comes to expect the system to bring her the salient facts about a document (whether these are composed of important sentences, or matched existing tags.)

While this may represent a new paradigm for scanning and tagging documents, we are also cautious about the implications. Lazy annotators may be tempted to accept them without critically assessing them, resulting in the propagation of poor tags. Within an educational context, one possibility would be to keep tag suggestions and automatic text highlighting at an imperfect level, to maintain students’ vigilance.

6.4 Interface bias towards ‘matched’ tags

Let us now consider the ‘matched tags’ recommender. Matched tags (exact matches, as with most social tagging tools) were privileged in the user interface over non-matched ones: the former were visible via the activation of a filter and highlighted in bright yellow zones directly in context within the document, while the latter were ‘hidden’ in the separate history window. Matched tag highlighting gives immediate feedback to annotators, and the satisfaction of seeing one’s tags highlighted on the text is akin to that gained in social bookmarking when one’s tagged pages show up with the rest of the world’s.

However, we again raise the question of the quality of tagging, whereby the emphasis could shift from reflectively submitting new tags, to submitting ‘visible’ tags (that is, matched by the dedicated recommender). Better presentation options must be devised, including a mechanism to display ‘non-matched’ tags in the main window. Although this has not been verified, it may be that the user interface design led annotators to forget that there might be other tags: it certainly did not actively remind them. This may have led them to submit more ‘copied-and-pasted’ tags. This added focus on the visual salience of highlighted text spans may also mean that matched tags became a way to cover the document with tags. By doing this, annotators received implicit feedback that they had read the document.

7 RELATED WORK

Our work is one strand in research on computational modelling of argumentation (e.g. COMMA [6]), but while other work focuses on the formalization of human or agent argument structures and processes, we place more emphasis on interaction design, and on the development of software

tools that forge a link between argumentation and current Web annotation tools and practices [11].

Our work builds on research into readers’ annotation practices, in which annotation is a means to record personal ideas and interpretations, including connections to additional scholarly documents, reformulations of the authors’ arguments, assessment of its significance or ‘warning’ signals to indicate key passages [14]. However, we are exploring the representational and interactional requirements for tools to enable these personal perspectives to be made public as a semiformal network that can be managed, extended, and contested. Current annotation tools [16] provide no support to manage what might be thought of as large scale annotations on annotations.

Ontology-based annotation tools are being developed as an essential part of the Semantic Web movement. However, these applications may in fact be better characterised as the supporting the ‘translation’ of information in the document into ontological entities. Although there may be debate about how to map an entity into an ontology, the material itself is not normally the focus of contention (such as the names of people, events, locations, processes). The tools certainly do not aim to support debate about the significance or meaning of an entity in a document.

Our use of recommendation filters derives from work on the summarisation of scientific papers. Potentially relevant passages can be delimited with multiple approaches, based on (i) the structure of the (scholarly) document [1] (ii) surface-based features [11] (iii) topical coherence [17] and (iv) rhetorical coherence measures [22]. Other work on literature-wide analysis on which we could draw includes identification of relevant documents by analysing their citations sections [9, 12]. Pivotal points can also be proposed to filter a network of documents and retain only the most important ones [5]. Nanba et al. [15] also propose an approach to both identify reference areas and the role [26] played by these areas. They consider the following roles: references indicating other researchers’ theories or methods used as a basis, references to related works to mention a contrast or a problem and other references.

Since researchers clearly need to annotate domain terminology, Semantic Web annotation tools are part of the solution. In CREAM [8], an annotation by mark-up mode is provided, enabling the user to select any piece of relevant information from the page and drag and drop it to create or instantiate the selected concept instance (researcher name, address...) Text fragments are extracted from the page to foster a semi-automatic annotation: the knowledge expert agent only has to validate the extracted elements.

However, following the social tagging paradigm, annotators in our approach will tag only those elements in a text that reflect their interests (there is no gold standard set of tags that can be automatically extracted, since there is no single, authoritative meaning). As we have argued on theoretical grounds elsewhere, the representational requirements for modelling discourse are different [13]. This

work is therefore better framed not so much as Semantic Web (controlling interpretation through consensus domain models) than as Pragmatic Web (foregrounding context, argument, interpretation and perspective) [3].

8 CONCLUSIONS AND FUTURE WORK

We offer this analysis as an example of a human-centred design process for collaborative knowledge structuring environments. We hope that the particular approach we are developing contributes to wider efforts to add greater representational expressiveness to social tagging, without in the process straitjacketing it.

Social bookmarking via freeform ‘folksonomic’ tagging is demonstrating its huge potential for collective indexing of materials through emergent vocabularies. In our approach, we have preserved the freedom that folksonomic tagging permits in what counts as a ‘tag’, added the option to classify tags, and introduced the option to link tags using familiar ‘research moves’, but predefined in order to leverage automated filtering and search. The ClaimSpotter prototype supports the collaborative annotation of documents using this representational scheme. We have summarised a detailed analysis of how annotators made use of the tool in their first hour of usage, describing the results under the themes of *Formalization*, *User Strategy* and *Interaction Design*.

This work is being developed in several directions. There is clearly scope to improve the interface design, and to add the kinds of flexibility that we see in social tagging interfaces such as recording tags as private, personalising recommendation filters, and enabling richer user profiles. ClaimSpotter is one of a suite of tools being developed in the *Hypermedia Discourse* project³ in which we are now developing a server to provide coherence relations-based tagging services, which we conceive as a form of *web pragmatics*.⁴

We are also testing the generality of the approach outside scholarly discourse, exploring the use of recommendation filters and discourse links in the *Laboranova* project⁵ which is focussing on the early stages of innovation when ideas are developed, debated, improved and evaluated. We are exploring the possibilities of introducing stimulus agents and serious games to strengthen proposals for innovation development by suggesting argumentative connections between ideas, supporting examples, diagnostic tools outputs or relevant experts.

9 ACKNOWLEDGMENTS

We are grateful to the reviewers for their helpful feedback. This research was supported by the Advanced Knowledge

Technologies (AKT) project, an Interdisciplinary Research Collaboration (IRC) sponsored by the UK Engineering and Physical Sciences Research Council (GR/N15764/01). The AKT IRC comprised the Universities of Aberdeen, Edinburgh, Sheffield, Southampton and The Open University.

10 REFERENCES

- [1] P. Bishop. Digital Libraries and Knowledge Disaggregation: the Use of Journal Article Components. In *Proceedings of the 3rd International Conference on Digital Libraries*. ACM, 1998.
- [2] R. J. Boland and R. V. Tenkasi. Perspective Making and Perspective Taking in Communities of Knowing. *Organization Science*, 6: 350–372, July 1995.
- [3] S. Buckingham Shum. Sensemaking on the Pragmatic Web: a Hypermedia Discourse Perspective. In *Proceedings of the 1st International Conference on the Pragmatic Web*. GI Lecture Notes in Informatics, September 2006.
- [4] S. Buckingham Shum, V. Uren, G. Li, B. Sereno, and C. Mancini. Computational Modelling of Naturalistic Argumentation in Research Literatures: Representation and Interaction Design Issues. *International Journal of Intelligent Systems*, 22(1): 17–47, 2006.
- [5] C. Chen. The Centrality of Pivotal Points in the Evolution of Scientific Networks. In *Proc. Int. Conf. Intelligent User Interfaces*, pages 98–105, ACM.
- [6] COMMA: 1st Int. Conf. on Computational Modelling of Argumentation, (Sept.’06), Liverpool, UK, IOS Press
- [7] B. G. Glaser and A. Strauss. *Discovery of Grounded Theory. Strategies for Qualitative Research*. Sociology Press, 1967.
- [8] S. Handschuh and S. Staab. Authoring and Annotation of Web Pages in CREAM. In *Proc. WWW2002: 11th Int. World Wide Web Conference*, 2002.
- [9] S. Hitchcock, L. Carr, Z. Jiao, D. Bergmark, W. Hall, C. Lagoze, and S. Harnad. Developing Services for Open Eprint Archives: Globalisation, Integration and the Impact of Links. In *Proceedings of the 5th Int. Conference on Digital Libraries*. ACM, 2000.
- [10] C. M. Hoadley and M. C. Linn. Teaching Science Through Online, Peer Discussions: SpeakEasy. In *The Knowledge* January 2005.
- [11] J. Kupiec, J. Pedersen, and F. Chen. A Trainable Document Summarizer. In *Proceedings of the ACM SIGIR’95 Conference*, pages 68–73. ACM, 1995.
- [12] S. Lawrence, C. L. Giles, and K. Bollacker. Digital Libraries and Autonomous Citation Indexing. *IEEE Computer*, 32(6): 67–71, 1999.

³ <http://kmi.open.ac.uk/projects/hyperdiscourse>

⁴ <http://www.pragmaticweb.info>

⁵ <http://www.laboranova.com>

- [13] C. Mancini and S. Buckingham Shum. Modelling Discourse in Contested Domains: a Semiotic and Cognitive Framework. *International Journal of Human Computer Studies*, 64(11): 1154–1171, 2006.
- [14] C. C. Marshall. Annotation: from Paper Books to the Digital Library. In *Proceedings of the 2nd ACM International Conference on Digital Libraries*, pages 131–140, Philadelphia, PA, USA, 1997. ACM.
- [15] H. Nanba and M. Okumura. Towards Multi-Paper Summarization using Reference Information. In *Proceedings of the IJCAI'99 Conference*, pages 926–931, 1999.
- [16] I. Ovsianikov, M. A. Arbib, and T. H. McNeill. Annotation Technology. *International Journal of Human Computer Studies*, 50(4): 329–362, 1999.
- [17] G. Salton, A. Singhal, C. Buckley, and M. Mitra. Automatic Text Decomposition Using Text Segments and Text Themes. In *UK Conference on Hypertext*, pages 53–65, 1996.
- [18] B. Sereno. A Document-Centric Semantic Annotation Environment to Support Sense-Making. PhD thesis (also available as *Technical Report KMI-06-13*), Knowledge Media Institute, The Open University, Milton Keynes, UK, September 2005.
- [19] B. Sereno, S. Buckingham Shum, and E. Motta. ClaimSpotter: an Environment to Support Sensemaking with Knowledge Triples. *Proc. Int. Conf. Intelligent User Interfaces*, pages 199–206, ACM.
- [20] F. M. Shipman and R. McCall. Supporting Knowledge Base Evolution with Incremental Formalization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 285–291. ACM, April 1994.
- [21] J. M. Swales. *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press, 1990.
- [22] S. Teufel and M. Moens. Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status. *Computational Linguistics*, 28(4):409–445, December 2002.
- [23] V. Uren, S. Buckingham Shum, G. Li, and M. Bachler. Sensemaking Tools for Understanding Research Literatures: Design, Implementation and User Evaluation. *International Journal of Human Computer Studies*, 64(5):420–445, 2006.
- [24] V. Uren, Buckingham Shum, S., Li, G., Domingue, J. and Motta, E. Scholarly Publishing and Argument in Hyperspace. *Proc. WWW2003: 12th Int. World Wide Web Conference*, May 20-24, 2003, Budapest.
- [25] K. Weick. *Sensemaking in Organizations*. 1995, Thousand Oaks, CA: Sage Publications.
- [26] M. Weinstock. Citation Indexes. In *Encyclopedia of Library and Information Science*, volume 5, pages 16–40. 1971.