Tracking Epistemic Beliefs and Sensemaking in Collaborative Information Retrieval

Simon Knight[1], Simon Buckingham Shum[1] and Karen Littleton[2]
[1] Knowledge Media Institute
[2] Centre for Research in Education & Educational Technology
The Open University, Milton Keynes, MK7 6AA, UK
+44 1908 654672
simon.knight@open.ac.uk

ABSTRACT
Collaborative information retrieval is an area of increasing interest. However, the wider sensemaking context in which it occurs is understudied. Furthermore, the role of users' beliefs about the domain they are querying, its structure, stability, complexity, and justifications – their epistemic beliefs – has been little studied in either individual or collaborative IR software development contexts. Here we discuss a tool to combine the knowledge mapping tool Cohere, with reference management capabilities through which the iterative, epistemically germane, potentially (but not necessarily) collaborative IR process may be 'brought out' both for sensemaking, and research purposes.

Author Keywords
Learning analytics; epistemic beliefs; collaborative information retrieval; information retrieval; sensemaking

ACM Classification Keywords
K.3.1 [**Computers and Education**]: Computer Uses in Education – *collaborative learning.*
General
Measurement, Documentation, Design, Human Factors,

INTRODUCTION
In this paper we discuss the role of epistemic beliefs – beliefs about the certainty, simplicity, source and justification for knowing [9:69] – in collaborative sensemaking for information retrieval (IR) tasks. This discussion is important because, we argue, the process of collaborative IR necessarily involves epistemic sensemaking around information needs; users making decisions about the sorts of knowledge they need, whether they have obtained this knowledge, and how to deal with it.

Collaborative IR is an increasingly common work and leisure activity [11,12]. People use search engines to research a plethora of topics from the academic search, to seeking products, holiday destinations, health information, and so on. These sorts of activities are often classed as 'exploratory' in nature; the user is not seeking one particular answer but rather to understand a domain, to 'sensemake' on it, and build a 'picture' of its structure.

Classic models of search indicate that the IR process involves: the identification of a need; the search to meet that need; the evaluation of results towards the need. This process has parallels in models of 'epistemic beliefs' – beliefs about the certainty, simplicity, source and justification for knowing [9:69]. In the educational context, recent evidence suggests that in the context of search engine IR, students spontaneously reflect about knowledge, and knowing [8] indicating that, "epistemological beliefs are a lens for a learner's views on what it is to be learnt" [2:8]. One model [13] has thus considered epistemic beliefs and self-regulation simultaneously; with a process from: task definition; to goal setting; enactment; and finally evaluation – a process which may be readily mapped to IR. It thus appears that sensemaking engages epistemic beliefs.

However, while there is a growing interest in collaborative IR [16], and a number of tools exist to support this activity, there is less exploration of how these tools might support the wider information processing and sensemaking process [14,18]. However, collaboration may have benefits for sensemaking in the context of IR – including in distributed settings [4]. For example, more 'expert' users in a given domain may open more results, and rate those results more relevant [19]; by making this process explicit, their *reasoning* and sensemaking may be relayed to the non-expert user.

Educational Workplace Context
These issues are of particular concern in educational contexts given the benefits of collaborative dialogue (see e.g. [7]), and the role that epistemic beliefs play in student knowledge management and information processing (see above). However, these concerns extend beyond the educational arena, and indeed even beyond that of 'workplace learning' – although epistemic beliefs play a role here too [5]. Any workplace practice involving the management, use, seeking, citation, and writing of, multiple documents will necessarily involve users in more or less explicit judgments regarding the usefulness of information

for their present task. Many if not most of these activities will involve collaborators, and many if not most could benefit from better systems to support their collaborative facilitation.

Furthermore, as distributed sensemaking research [4] has noted, presently once information has been sought, found, and structured by a user, it is often lost. This is true within workplaces and more broadly. The concern is not that users should more easily find the "correct answer" to some problem as instantiated through a search query. Rather, the concern regards the loss of the sensemaking process of structuring a domain, relating concepts, and meaning making between that structuring and the 'task at hand' – the making of epistemic judgments.

Within the academic and educational workplace the IR process is particularly used for reference management practices. The typical workflow will involve:

1. Identifying a problem
2. Searching the literature
3. Saving relevant documents (alongside bibliographic metadata)
4. Structuring results; sensemaking around them
5. Writing a document, and publishing it.

Parallels can be drawn here between this process, and those of the classic models of IR. Moreover, while this process – and reference management tools (such as the Open Source Zotero) are broadly academic, the process of finding documents, judging relevance, and writing about them (while ensuring appropriate citation and attribution) is a common workplace practice, and indeed bibliographic tools are setup to facilitate the saving of metadata regarding many document types – not just academic.

The Problem
The problem, then, is how to facilitate the collaborative (possibly distributed) sensemaking of users engaging in exploratory IR while maintaining key workplace – particularly academic – practices such as citation management, and publishing. Addressing this issue should also hold other benefits. For example:

1. In introducing users to 'new' domains – by making the sensemaking process explicit and visual, such data can be explored at later times.
2. In understanding the relationship between user's 'published work', and the process they have gone through to create these documents
3. In understanding how websites and topics stand in relation to each other (by inference from user's notes on those websites)
4. In understanding the user's learning process – as an educational technologist, this is my particular interest.

SOLUTIONS AND RELATED WORK
A number of systems have been designed to meet some of the needs addressed here. Those needs in particular are:

1. The need for a shared search, with shared awareness regarding queries made
2. The need for a shared document space
3. The two related needs for:
   a. a shared understanding of the domain structure; how the user is 'sensemaking' on the information, and,
   b. this shared sensemaking to be discursive in nature
4. The need for a smooth workflow from search, to publishing.

The following four sections discuss some of these tools, marking with parenthesised numbers where a particular need (above) is met. Space does not permit a comprehensive review of tools, however the tools selected highlight particular means to address the needs, and have been drawn on in the development of our own tool set, as we discuss in the following section.

Coagmento
Coagmento [15] is a browser addon which provides users with a sidebar chat and (shared) search history function. As well as these two tools, Coagmento provides a shared document space (for sharing files) and a collaborative writing tool (etherpad). It thus ostensibly meets 1-4 above. However, although searches may be saved and marked, the shared sensemaking (3) is not structured for argumentation or sensemaking on complex issues. Similarly, while Coagmento provides some tools to facilitate publishing (4), it does not integrate with other tools – including citation management – which may be important for many users.

Search Together (with CoSense)
CoSense [14,17] was designed as a tool to be used in tandem with Search Together [10]. In combination, those tools provide a shared search history and chat features, alongside the ability to make notes on webpages. All this information can then be filtered and viewed in a number of ways. While this approach was successful in enhancing the sensemaking (3) experience, for work involving argumentation and reasoning, more structured environments may be preferable. In addition, it does not provide as dynamic a space for shared documents as Coagmento, nor the same functions for 'writing up' and publishing (including citation management) as we propose.

Docear
A completely different type of tool is Docear [1]. Docear is an academic literature suite with a built in mind-mapping tool (Freeplane) which allows users to mind-map both their references and notes made on them (which are created as separate nodes), and easily copy these nodes (with references) into Microsoft Word for document authoring and publishing. It thus facilitates an *individual* sensemaking experience rather than collaborative (3a), although maps can be shared, they cannot be easily co-edited. It also facilitates the storing and 'flow' of document creation (4),

although it does not record queries made, or provide shared space for such information.

Cohere

In contrast to Docear, Cohere [3] is designed to facilitate collective sensemaking (3) via a social web annotation and bookmarking tool, with a user-customizable visual language, with the ability to make user-defined meaningful connections between annotations, to generate a range of visualizations. It has been successfully used with discourse analytic tools to explore student's argumentation skills [6]. It allows users to annotate documents (with anchored annotations) and organize these with user generated nodes to create knowledge maps in which nodes are 'idea types' and connections give a semantic relation between nodes (at most basic, pro/con/neutral).

Summary of Existing Tools

While there are tools to collaborate over search, they do not integrate with bibliographic management, knowledge mapping, or document authoring and management tools. Current tools which facilitate collaborative bibliographic management (particularly Zotero and Mendeley) do not well support IR, online collaborative authoring, or knowledge mapping. Tools which do support collaborative authoring (ether pad in Coagmento, google docs, etc.) do not function well with bibliographic data, search, or knowledge mapping, and similarly those which support knowledge mapping – individually (Docear) or collaboratively (Cohere) have limitations as discussed above. The proposed tool would cover 1-4 above, and allow implicit collaboration at later dates, allowing for a contested collective intelligence which is dynamic, structured, and long lasting in nature.

CONCEPT OUTLINE

Founded in this prior work, in particular the Docear concept, we are building a tool to facilitate collaborative IR within the context of a document authoring, and citation management, system. It will be, to our knowledge, the only such tool set implemented in whole. Furthermore, the interoperability of various open source solutions has benefits for end users, the developer community, and researchers. In this example, we envisage bibliographic management via Zotero, shared sensemaking (knowledge mapping and document annotation) through Cohere, and document authoring through WordPress. The Zotero and Cohere APIs facilitate interoperability, while ZotPress – a WordPress plugin – allows citation management in a number of formats through the use of a Zotero library[1].

---

[1] ZotPress does not provide as much functionality as the Zotero word addin, for example, the user is still required to type a code of the form [zotpress InText item="xyz"] as opposed to simply selecting the citation from their library to 'insert' intext. However, it is one of very few reference management tools with online collaborative capabilities and is still under development.

While Zotero is primarily aimed at academics, it is capable of saving and indexing any media type, including saving the full copy (and backing these up to a WebDav server). It now works on a variety of browsers, although it was built for Firefox. In our proposal, new references should be associated with the search query made to find them, and 'connected' to that node on the Cohere map.

Issue	Solution
The need for a shared search, with shared awareness regarding queries made	Queries logged as Cohere nodes. Documents saved in Zotero marked as 'saved' on the results page. Such document nodes are also associated with the 'query node' form which they were saved.
The need for a shared document space	Raw documents through shared Zotero space. Document authoring through WordPress export from Cohere (with ZotPress to maintain references).
Shared understanding & sensemaking of the domain structure	Cohere maps to support collaborative, distributed, asynchronous sensemaking. Nodes – from documents and user created – may be 'connected' to queries, and any other public (or user) node.
Shared discursive sensemaking	Annotation and semantic markup (through nodes and connections). Possibility for chat integration. Shared document space.

Table 1. Mapping issues and solutions in sensemaking for IR

Epistemic Beliefs

A key benefit of the inclusion of Cohere in this system is in the making explicit of epistemic assumptions (in a broad sense) through the Cohere structuring as nodes of: queries; results; annotations; and general notes. This will create a structured map of the domain which may be added to by subsequent searchers. In education, this provides opportunity to encourage users to add different sorts of node, or connection – exploring methods (node type), or pros/cons depending on what connection types dominate their maps. Beyond education, these maps also provide an explicitly structured means through which to understand the prior searchers sensemaking on the domain, as structured around a set of searches made, and – potentially – document produced through WordPress.

CONCLUSION

By integrating a number of open source tools – Zotero and WordPress (using ZotPress), and Cohere – with a number of new features, which are common to collaborative IR tools, a new knowledge management tool can be implemented. The tool described seeks to make explicit the sensemaking process from problem identification, IR, document processing, to authoring, publishing, and review. The benefits for workplace practice – particularly academia

and education generally – are clear. Further benefits are directly related to the 'bringing out' of the sensemaking process; its reification through technological artefacts allows the exploration of those artefacts in particular via the use of Learning Analytics. Such analysis is one – salient for me – example of how this type of exploration may be particularly useful for algorithmic mediation, with other features based on suggested search, or automatic structuring, also implementable in line with prior work. This approach is a novel one, and it should be noted that we see the combined tool set as greater than the sum of its parts. We claim that in making open the process of sensemaking, and combining tools to create a smooth toolset across which this sensemaking may occur, greater insight can be gained both by users, and analysts (including educationalists). The claim is further, that these tools may *mediate* the collaborative IR process. By this we mean two things, firstly that it may bring out salient factors in the process making these explicit for automated and human analysis. Secondly, that the toolset itself in becoming part of the process, influences how that process is conducted and encourages users to engage in particular sorts of activity – including the collaborative sensemaking described.

While some work has explored the distributed sensemaking process, the proposal here focuses in on a particularly salient facet of regulating the IR process. This approach implicates epistemic beliefs as the lens through which users see, and address, problems. Presently this information is lost – new searchers must start from their default lens. The outline provides a means through which to track users epistemic shifts – the changes, combinations, and conflicts in lenses used as implicit theories in the action of IR for some problem domain.

REFERENCES
1. Beel, J., Gipp, B., Langer, S., and Genzmehr, M. Docear: An academic literature suite for searching, organizing and creating academic literature. *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries*, (2011), 465–466.
2. Bromme, R., Pieschl, S., and Stahl, E. Epistemological beliefs are standards for adaptive learning: a functional theory about epistemological beliefs and metacognition. *Metacognition and Learning 5*, 1 (2009), 7–26.
3. Buckingham Shum, S. Cohere: Towards web 2.0 argumentation. *Frontiers in Artificial Intelligence and Applications*, IOS Press (2008), 97.
4. Fisher, K., Counts, S., and Kittur, A. Distributed sensemaking: improving sensemaking by leveraging the efforts of previous users. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, (2012), 247–256.
5. Harteis, C., Gruber, H., and Hertramph, H. How epistemic beliefs influence e-learning in daily work-life. *Educational Technology & Society 13*, 3 (2010), 201–211.
6. De Liddo, A., Buckingham Shum, S., Quinto, I., Bachler, M., and Cannavacciuolo, L. Discourse-centric learning analytics. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, ACM (2011), 23–33.
7. Littleton, K. and Howe, C. *Educational dialogues: understanding and promoting productive interaction*. Routledge, Abingdon, Oxon, 2010.
8. Mason, L., Ariasi, N., and Boldrin, A. Epistemic beliefs in action: Spontaneous reflections about knowledge and knowing during online information searching and their influence on learning. *Learning and Instruction 21*, 1 (2011), 137–151.
9. Mason, L., Boldrin, A., and Ariasi, N. Epistemic metacognition in context: evaluating and learning online information. *Metacognition and Learning 5*, 1 (2009), 67–90.
10. Morris, M.R. and Horvitz, E. SearchTogether. ACM Press (2007), 3.
11. Morris, M.R. A survey of collaborative web search practices. ACM Press (2008), 1657.
12. Morris, M.R. Collaborative Search Revisted. (2013).
13. Muis, K.R. The Role of Epistemic Beliefs in Self-Regulated Learning. *Educational Psychologist 42*, 3 (2007), 173–190.
14. Paul, S.A. and Morris, M.R. CoSense: enhancing sensemaking for collaborative web search. *Proceedings of the 27th international conference on Human factors in computing systems*, (2009), 1771–1780.
15. Shah, C. and González-Ibáñez, R. Exploring information seeking processes in collaborative search tasks. *Proceedings of the American Society for Information Science and Technology 47*, 1 (2010), 1–7.
16. Shah, C. Collaborative Information Seeking: A Literature Review. In A. Woodsworth, ed., *Advances in Librarianship*. Emerald Group Publishing, Bingley, 2010, 3–33.
17. Sharoda, P. and Morris, M.R. CoSense. ACM Press (2009), 1771.
18. Sharoda, P. and Morris, M.R. Sensemaking in Collaborative Web Search. *Human-Computer Interaction 26*, (2011), 72–122.
19. Zhang, X., Liu, J., Yuan, X., Cole, M., Belkin, N., and Liu, C. Do users with different domain knowledge select different sets of documents? *Proceedings of HCIR 2012*, (2012).