Designing a Citation-Sensitive Research Tool: An Initial Study of Browsing-Specific Information Needs

Stephen Wan[†], **Cécile Paris**[†],
[†] ICT Centre,
CSIRO, Australia
Firstname.Lastname@csiro.au

Michael Muthukrishna[†], **Robert Dale**[‡]
[‡]Centre for Language Technology
Faculty of Science
Macquarie University, Australia
rdale@science.mq.edu.au

Abstract

Practitioners and researchers need to stay up-to-date with the latest advances in their fields, but the constant growth in the amount of literature available makes this task increasingly difficult. We investigated *the literature browsing task* via a user requirements analysis, and identified the information needs that biomedical researchers commonly encounter in this application scenario. Our analysis reveals that a number of literature-based research tasks are preformed which can be served by both generic and contextually tailored preview summaries. Based on this study, we describe the design of an implemented literature browsing support tool which helps readers of scientific literature decide whether or not to pursue and read a cited document. We present findings from a preliminary user evaluation, suggesting that our prototype helps users make relevance judgements about cited documents.

1 Introduction

Practitioners and researchers in all fields face a great challenge in attempting to keep up-to-date with the literature relevant to their work. In this context, search engines provide a useful tool for information discovery; but search is just one modality for gathering information. We also regularly read through documents and expect to find additional relevant information in referenced (cited or hyperlinked) documents. This results in a browsing-based activity, where we explore connections through related documents.

This browsing behaviour is increasingly supported today as publishers of scientific material deliver hyperlinked documents via a variety of media including Adobe's Portable Document Format (PDF) as well as the more conventional web hypertext format. Given appropriate document databases and knowledge of referencing conventions, it is relatively straightforward to support the automatic downloading of cited documents: such functionality already exists within reference managers such as *JabRef*[1] and *Sente*[2]. This 'blind downloading', however, does not address the question of the relevancy of the linked document for the reader at the time of reading. Apart from the publication details of the reference and the citation context, readers are provided with very little information on the basis of which to determine whether the cited document is worth exploring more thoroughly. Given the potentially large number of citations that may be encountered, this results in the following *browsing-specific scenario*: how can we help a user quickly determine whether the cited document is indeed worth downloading, perhaps paying for, and reading?

In the study presented here, we focussed on the needs of biomedical researchers, who are often time-poor and yet apparently spend 18% of their time gathering and reviewing information (Hersh, 2008). They regularly search through repositories of online scholarly literature to update their expert knowledge; in this domain, the penalty for not staying up-to-date with the latest advances can be severe, potentially affecting medical experiments. In our work, we found that two thirds of researchers regularly engaged in browsing scientific literature. Given the prevalent use of the browsing modality, we believe that novel research tools are needed to help readers make decisions about the relevance of cited material.

To better understand the user's information needs that arise when reading and browsing through academic literature, and to ascertain what NLP techniques we might be able to use to help support them, we conducted a user require-

[1] jabref.sourceforge.net
[2] www.thirdstreetsoftware.com

Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries, ACL-IJCNLP 2009, pages 45–53,
Suntec, Singapore, 7 August 2009.

ments analysis. It revealed a number of common problems faced by readers of scientific literature. These served to focus our efforts in designing and implementing a browsing support tool for scientific literature, referred to here as CSIBS.

CSIBS helps readers decide which cited documents to read by providing them with information which is useful at the point when citations are encountered. The application provides information about the cited document and identifies important sentences in that document, based on the user's current reading context. The key observation here is that the reading context can indicate *why* the reader might be interested in the cited document. In addition to meta-data about the cited document, and its abstract, a *contextualised preview* is shown within the same browser in which the citing document is being viewed (for example, Adobe Acrobat Reader or a web browser), thus avoiding an interruption to the user's primary reading activity. This contextualised preview contains important sentences from the cited document that are related to the reading context.

We present related work on understanding information needs in Section 2; we outline our user requirements analysis in the domain of scientific literature in Section 3; and the results of the analysis and our understanding of the browsing-specific information needs are presented in Section 4. In Section 5, we describe a tool developed to meet the most pressing of these information needs. Section 6 presents a feedback from an initial evaluation. We conclude by discussing our overall findings in Section 7.

2 Related Work

2.1 Information Needs

Existing work on information needs, beginning with Taylor (1962), typically focuses on mapping from a particular query to the underlying interest of the user. In a recent example of such work, Henrich and Luedecke (2007) describes methods for constructing lists of domain-specific key words which may correspond well to user interests. However, we are interested in relating information needs to user tasks in scenarios in which there is no explicit query, as in Bystrm et al. (1995); in particular, our work focuses on browsing scenarios. Toms (2000) presents a study of browsing behaviour over electronic texts and examines the differences between searching and browsing. In that work, browsing is performed across multiple news articles where the links between articles are inferred based on topic similarity. In contrast, we consider explicit hypertext links which are linguistically embedded in the document as citations, where the embedding text serves as link anchors.

2.2 Information Needs in Biomedicine

Ely et al. (2000) present an overview of the information needs of practicing clinicians, deriving a set of commonly asked questions. Although we are interested in doctors as users, the type of information needs presented in this paper relate to the activity of conducting scientific investigation, rather than that of treating a patient.

Task-based analyses of the biomedical domain have been studied by Bartlett and Neugebauer (2008) and Tran et al. (2004). Their analyses, like ours, are task-based and use qualitative studies to uncover the underlying uses of information. However, the tasks outlined in these related works are focused on a specific set of information needs in a research area: for example, the determination of a functional analysis of gene sequences. Our work differs in that we wish to take a more general view in order to elicit information needs to do with scientific research, at least at the level of biomedical sciences.

The information needs and tasks of academic users have been studied previously by Belkin (1994), who focuses on scholarly publications in the humanities domain. We perform an investigation along similar lines, but with a focus on academic literature used to conduct scientific research.

2.3 Using Scientific Literature

The genre of academic literature, and the development of technologies to support researchers as users, has been studied by several groups working in automatic text summarisation. Teufel and Moens (2002) describe a summarisation approach that extracts text from documents and highlights the rhetorical role that an extract plays within the originating document (for example, stating the *Aim* of an experiment). Qazvinian and Radev (2008) present an approach to summarising academic documents based on finding citation contexts in the entire set of published literature for the document in question. Both approaches, however, treat the cited document in isolation of the read-

ing context and do not actively support the reading task.

3 Understanding How Researchers Browse through Scientific Literature

To determine what readers of scientific literature want to know about cited documents, we conducted a user requirements analysis. Our method is based on Grounded Theory (Glaser and Strauss, 1967), a commonly used approach in Human Computer Interaction (Corbin and Strauss, 2008). We began by interviewing subjects from an appropriate user demographic and recording their verbal descriptions about a real scenario situated in their day-to-day activities. Following this, we designed a questionnaire for wider participation which presented scenario-based questions attempting to uncover their information needs and tasks. Participants were asked to provide free text answers. The responses were then collated and analysed for commonalities, bringing to the fore those issues that were salient across the participants. We report on the questionnaire design and responses in this paper.

Beginning with such a study can reduce the risk of building tools that have only limited utility. This is particularly true of new and less understood application scenarios, such as the one explored here.

3.1 Questionnaire Design

An online questionnaire was used to reach participants who actively read academic literature.[3] To encourage participation, the questionnaire was limited to 10 questions, which were formulated independently of any particular scientific domain.

We were explicit about the aims of the questionnaire by providing an initial brief, stating that the feedback from participants would be used to develop new tools for browsing through scientific literature. Within the questionnaire, to prepare participants for our scenario-based questions, the first few questions were basic and concerned the general usage of scientific literature. For example, we asked about the high-level reasons for which they used scientific literature (e.g., 'To learn about a new topic'; 'To update your knowledge on a particular topic'). Participants could also specify their own reasons. In addition, we also asked them about the frequency of their literature browsing activity.

The main section of the questionnaire consisted of a series of questions, corresponding to the issues we wanted to explore:

1. What information needs do researchers have of a cited document, and what specific tasks does this information serve?
2. What makes it difficult for researchers to find the answers to their questions about cited documents?
3. What tasks are potential targets for automation?

Questions were to be answered with free text responses, focussed by presenting a scenario in which the researcher encounters a citation whilst reading a scientific publication. The first question above aims to better understand the researchers' information needs and tasks; the second and third are concerned with ideas for potential applications which could benefit from NLP and IR research.

To address the first research issue, participants were asked to recall a recent experience in which, while reading a publication, they had encountered a citation. Within this context, participants were asked to describe what questions they may have had of the cited document. To clarify how these questions relate to a specific context of use, respondents were then asked to relate the questions they identified back to some task undertaken as part of their research work.

Responses regarding the difficulties encountered in satisfying information needs were collected with respect to the participants earlier responses. So as to not bias the participant, the question was phrased neutrally. We asked what aspects of scientific literature and current technology made it easy or hard to find answers to the participants' personal research questions. We examined responses with the aim of determining how technology might reduce the burden of knowledge discovery. Responses were again focused by using the same scenario as in the previous question.

The third research issue was explored via two separate questions. The first presented the participants with a scenario in which they had access to a non-expert human assistant who could perform one or more simple tasks identified in their earlier responses; they were then asked what kinds

[3] The online questionnaire tool, SurveyMonkey (`www.surveymonkey.com`), was used to implement the questionnaire as an online interactive form.

of tasks they would delegate to such an assistant. A second, more direct, question was presented requiring participants to describe which tools they would like to use, or to suggest new tools that would help them in the future, when it came to browsing through scientific literature.

Finally, optional questions about the participants' research backgrounds were presented at the end of the questionnaire. These were deliberately placed last to reduce barriers to completion.

4 Questionnaire Data Analysis

4.1 Analysing the Results

We recruited users with a background in biomedical life sciences since we had access to an extensive corpus of documents in this domain with which to build some kind of application. Note, however, that our questions were not specific to this domain, and the questionnaire could potentially be re-run with participants from a different scientific background.

We contacted 36 users who might be interested in life sciences publications. Of these, 24 participants started the questionnaire, and 18 completed it. Of the 24 participants, two thirds indicated that they browsed through academic literature at least once a week.

The written responses were separately analysed by three of the authors. Responses to each question were examined, checking for repeated terms and concepts that could form the basis of clustering. Salient information needs were matched to corresponding tasks, and commonly mentioned areas of difficulty and suggestions for delegation were grouped. Once each author had performed his or her own analysis, the salient groupings for each question were collaboratively determined, consolidating the three analyses performed in isolation. The most salient groupings were then examined for potential tasks that might be automated.

4.2 Questionnaire Data

We now present the results of the analysis. These are organised with respect to each of the three research issues.

4.2.1 Questions of the Cited Document

Figure 1 presents the most frequently indicated information needs and the most frequent tasks that were identified. The information needs can be grouped into two main categories. The first, which we refer to as *meta-data needs*, refers to information about the document external to the document content itself. These needs could be met by a series of database queries *about* the document, involving, for example, the author information and the citation counts for the document. We note that, often, the abstract can also be retrieved via a database query (and thus does not require any in-depth text analysis of the cited document), although technically this is not meta-data. In terms of the underlying task, this kind of generic information may be used in *deciding whether to trust the cited source*.

Information Needs	Freq
[md] About accessing the full text	9
[co] Article details (Definition, Methods, Results)	7
[md] About the authors	6
[md] About the publication date	5
[co] About relevance to own work	4
[md] The abstract	3
[co] The references	3

Participant Task	Freq
Deciding whether to believe the citation	4
Finding baselines for experiments	3
Comparing own ideas to article	3
Finding information to justify the citation	3
Finding information about methods	2
Finding additional references	2
Updating clinical knowledge	2
Conducting a survey of the literature	2
Identifying key researchers in the field	2
Updating research knowledge	2

Figure 1: Principal information needs and tasks of participants with regard to citations. In the first table, information needs are prefixed by 'md' for *meta-data* and 'co' for *content-oriented*. 'Freq' indicates the number of occurrences in the results.

The second category of information needs, which we refer to as being *content-oriented*, can be met by providing information sourced from *within* the cited document. This type of information facilitates multiple tasks. For example, these might include understanding why a document was cited, or finding new baselines to design new experiments. We refer to these tasks in general as *citation-focused*, as some underlying information need is triggered by the text that the participant has just read, whether this is for advancing one's understanding of a topic, or pursuing a specific line of scientific inquiry.

4.2.2 Difficulties in Finding Answers

This question required participants to voluntarily reflect on their own research practices, a process that is influenced partially by their expertise in research and their exposure to different research tools. Some responses described features of software that were appealing, while others related to the difficulties faced by researchers in finding relevant information. In this paper, we present only the subset of responses that concern the difficulties encountered, since this will influence the functionality of new research tools. These responses are presented in Figure 2.

Difficulties	Freq
Finding the exact text to justify the citation	3
Poor writing	2
Comparing documents	1
Resolving references to the same object	1

Figure 2: Difficulties in finding information.

In general, the difficulties concerned some kind of analysis of text. We note that these tasks are largely citation-focused, requiring content-oriented information. Examples of comments regarding this task are presented in Figure 3. For example, participants wanted to know how the cited document compared the citing document from the perspective of experimental design. However, the citation-focused task that was most commonly mentioned as difficult was that of justifying citations. Participants mentioned that reading through the entire cited document for this purpose was a tedious task, particularly when looking for information in poorly written documents.

4.2.3 Tasks for Automation

Our analysis of responses to the task automation questions revealed two interesting outcomes: delegation occurred often with the use of key words, and participants expressed the need for tools to express relationships between domain concepts. These are presented in Figure 4.

Responses to the question regarding task delegation revealed that for research-oriented tasks, participants felt the need to direct assistants through the use of key words. This is consistent to responses to earlier questions detailing what aspects of current technology were attractive, including user interface conventions such as key word highlighting. Otherwise, the other reported

Citation usually does not include the position of the information in the cited article . . . it might be necessary to read all of the article to find it in another reference and so on.

If the first report was only citing the second report for a small piece of information, that *information may be hard to locate* in the second report.

The original reference may have just cited a very small component of the second report, either just a comment made in the discussion or a supplemental figure . . . *It may take a while to locate and justify the citation* if it isn't the major finding of the report.

If I see a citation in a report that I am interested in, *I generally want to know if the cited report actually supports the statement in the original report*. Very often – way too often – citations do not. For all important citations I track down the original cited work and verify that it actually says what it is supposed to.

Figure 3: Some sample responses from users with regard to justifying citations; emphases added.

Automation Possibilities	Freq
Search cited document for key words	4
Search for further publications using key words	3
Refine search using related concepts	6

Figure 4: Potential candidates for a new research tool.

delegated task was that of simple database entry of publication records. We interpret these responses as indicating that participants are not overly willing to hand over responsibility for complex tasks to assistants. If delegation of more research-oriented activities occurs, participants want to understand how and why results were obtained. While responses were made assuming delegation to human assistants, we believe that such issues are even more crucial for results obtained via automated means.

Suggested novel features centered upon a better representation of relationships between domain concepts to be used for query refinement. Responses included expressions such as "refined search", a handling of user-specified "mind maps" (for repeated searches), and the use of "trails" explaining how results connected to search terms, key words and the author.

5 Prototype Requirements

As a result of these findings, we chose to build a tool that meets the two types of information needs revealed in the initial user requirements study. The

purpose of the resulting tool, CSIBS, is to help readers prioritise which cited documents are worth spending time to download and read further. In this way, CSIBS helps readers to browse and navigate through a dense network of cited documents.

To facilitate this task in accordance with the elicited user requirements, CSIBS produces an alternate version of a published article that has been prepared with pop-up previews of cited documents. Each preview contains meta-data, the abstract and content-oriented information. It is provided to the user to help perform research tasks that arise as a consequence of encountering a citation and needing to investigate further. The preview is not intended to serve as a surrogate for the cited document. Rather, it is aimed at helping readers make relevance judgements about citations.

The meta-data helps the user to appraise the citation and to make a value judgement about the work cited. The abstract provides a generic summary of the cited document, indicating the scope of the work cited. The content-oriented information supports any *citation-focused* tasks, for example citation justification, through the provision of detailed information sourced from within the cited document. We refer to this as a *Contextualised Preview*. It is constructed using automatic text summarisation techniques that tailor the resulting summary to the user's current interests, here approximately represented by the citation context: that is, the sentence in which the citation is linguistically embedded. We briefly describe CSIBS, in this section; for a full description, see Wan et al. (2009).

Each preview appears in a pop-up text box activated by moving the mouse over the citation. The specific interaction (a double click versus a "mouse-over") depends on whether the article is displayed via a web browser or as a PDF document. Figure 5 shows the resulting pop-up for the PDF display.

5.1 A Meta-Data Summary and Abstract

Participants often wanted a generic summary outlining the overall scope and contributions of the cited work. This is typically available via the abstract. Additionally, CSIBS presents a variety of meta-data returned from queries to an online publications database:[4]

- The full reference: This provides readers with the date of publication and the journal title, amongst other things.
- Author Information: CSIBS can include data to help the reader establish a level of trust in the citation, primarily focusing on information about the authors' affiliations and the number of related citations in the research area.
- The citation count for the cited document: Participants indicated that this was useful in appraising the cited article.

These pieces of information were commonly identified as useful in helping readers make value judgements about the cited work. This is perhaps an artifact of the biomedical domain, where research has a critical nature and concerns health and medical issues.

5.2 A Contextualised Preview

To generate the contextualised preview of the cited document, the system finds the set of sentences that relate to the citation context, employing approaches for summarising documents that exploit anchor text (Wan and Paris, 2008). Following Spark Jones (1998), we specify the *purpose* of the contextualised summary along particular dimensions, indicated here in italics:

- The *situation* is tied to a particular context of use: an in-browser summary triggered by a citation and its citing context.
- An *audience* of expert researchers is assumed.
- The intended *usage* of the summary is one of preview. We assume that the reader is making a relevance judgement as to whether or not to download (and, if necessary, buy) the cited document. Specifically, the information presented should help the reader determine the level of trust to place in the document, understand why the article is cited, and decide whether or not to read it.
- The summary is intended only to provide a partial *coverage* of the whole document, specifically focused on content that directly relates to the citation context.
- The style of the summary is intended to be *indicative*. That is, it should present specific

[4] www.embase.com

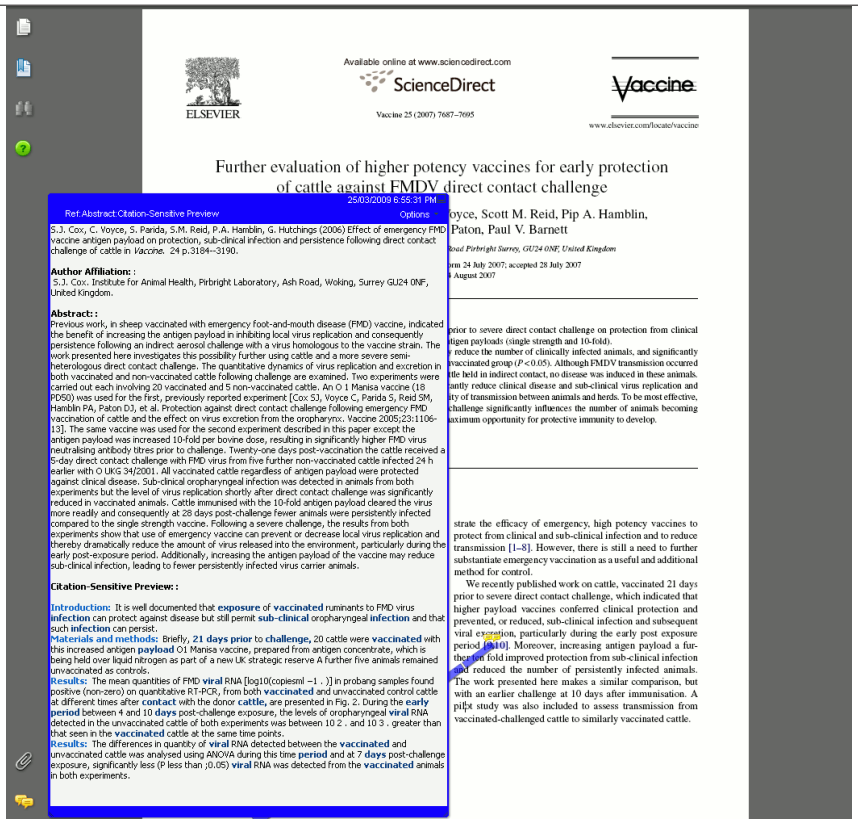

Figure 5: A sample pop-up with an automatically generated summary, triggered by a mouse action over the citation. Extracted sentences are grouped together by section titles. Words that match with the citation context are coloured and emboldened.

details to facilitate a relevance judgement, allowing the user to determine if the cited document can be used to source more information on a topic, as opposed to just mentioning it in passing.

To create the preview summary, the cited document is downloaded from a publisher's database[5] in its XML form and then segmented into sections, paragraphs and sentences. Each sentence in the cited document is compared with the citation context in order to find the best justification sentences for that particular citation. Due to the limited space available in the pop-up, the number of extracted sentences is capped at a predefined limit, currently set to four. Using vector space methods (Salton and McGill, 1983) weighted with term frequency (and omitting stop words), the best matching sentence is defined as the one scoring the highest on the cosine similarity metric with the citation context. The attractiveness of this approach lies in its simplicity, resulting in a fast computation of a preview ($\approx$ 0.03 seconds), making the process amenable to batch processing of multiple documents or, in the future, live generation of previews at runtime. To help with the readability of the resulting preview, the system also extracts structural information from the cited document. In particular, for each extracted sentence, the system identifies the section in which it belongs; the extracted sentences are then grouped by section, and presented with their section headings, as illustrated in Figure 5.

CSIBS focuses on returning precise results, so that the system does not exacerbate any existing information overload problems by burdening the reader with poorly matching sentences. To achieve this, we currently use exact matches to words in the citation context; in on-going work, we are exploring methods to relax this constraint without hurting performance. In line with our user requirements analysis, we have designed the tool so that the user is able to easily see how the summary was constructed. Matching tokens are highlighted, allowing the reader to understand why specific sen-

[5] `www.sciencedirect.com`

tences were extracted.

6 Initial Feedback

6.1 Evaluation Overview

We built a prototype version of CSIBS and conducted a preliminary qualitative evaluation. The goal was to examine how participants would react to the pop-up previews. The feedback allows us to further clarify our analysis and subsequent development.

We asked participants to view a number of pop-up previews in order to answer the question: *Is the Citation Justified?* This was one of the more difficult questions that researchers found challenging when making a relevancy judgement. The actual judgements are not important in this evaluation. Instead, we gauged the reported utility of the prototype based on the participants' self-reported confidence when performing the task. To capture this information, participants were asked to score their confidence on a 3-point Likert scale.

Three biomedical researchers, all of whom had taken part in our original user requirements analysis, participated in the evaluation. Each participant was shown nine different passages containing a citation context, each situated in a different *FEBS Letters*[6] publication (which was also presented in full to the participants). At each viewing of a citation context, two supporting texts were provided with which the participant was asked to answer the citation justification question. For all participants, the first supporting text was produced by a baseline system that simply provided the full reference of the citation. The second was either the abstract or the contextualised preview, which in this evaluation was limited to three sentences. Meta-data was not presented for this study as we specifically wanted feedback on the citation justification task.

The small sample size does not permit hypothesis testing. However, we are encouraged by the comparable positive gains in self-reported confidence scores (Abstract: +1.2 versus CSIBS: +2.2) compared to simply showing the full reference. Since both preview types were positive, we assume that these types of information facilitated the relevance judgements. Participants also reported that, for the contextualised preview, 2 out of 3 sentences were found to be useful on average.

[6]The journal of the Federation of Europeans Biochemical Societies.

The qualitative feedback also supported CSIBS. One participant made some particularly interesting observations regarding selected sentences and the structure of the cited document. Specifically, useful sentences tended to be located deeper in the cited document, for example in the methods sections This participant suggested that, for an expert user, showing sentences from the earlier sections of a publication was not useful; for the same reason, the abstract might be too general and not helpful in justifying a citation. Finally, this participant remarked that, in those situations where each document downloaded from a proprietary repository incurs a fee, the citation-sensitive previews would be very useful in deciding whether to download the document.

7 Conclusions

In this paper, we presented an analysis of browsing-specific information needs in the domain of scientific literature. In this context, users have information needs that are not realised as search queries; rather these remain implicit in the minds of users as they browse through hyperlinked documents. Our analysis sheds light on these information needs, and the tasks being performed in their pursuit, using a set of scenario-based questions.

The analysis revealed two tasks often performed by participants: the appraisal task and the citation-focused task. CSIBS was designed to support the underlying needs by providing meta-data information, the abstract, and a contextualised preview for each citation. The user requirement of search refinement was not directly addressed in this work, but could be met by techniques of query refinement in IR, synonym-based expansion in summarisation, and of course, additional user specified key terms. In future work, we will explore these possibilities. Our results to date are encouraging for the use of NLP techniques to support readers prioritise which cited documents to read when browsing through scientific literature.

Acknowledgments

We would like to thank all the participants who took part in our study. We would also like to thank Julien Blondeau and Ilya Anisimoff, who helped to implement the prototype.

References

Joan C. Bartlett and Tomasz Neugebauer. 2008. A task-based information retrieval interface to support bioinformatics analysis. In *IIiX '08: Proceedings of the second international symposium on Information interaction in context*, pages 97–101, New York, NY, USA. ACM.

Nicholas J. Belkin. 1994. Design principles for electronic textual resources: Investigating users and uses of scholarly information. In *Current Issues in Computational Linguistics: In Honour of Donald Walker.Kluwer*, pages 1–18. Kluwer.

Katriina Bystrm, Katriina Murtonen, Kalervo Jrvelin, Kalervo Jrvelin, and Kalervo Jrvelin. 1995. Task complexity affects information seeking and use. In *Information Processing and Management*, pages 191–213.

Juliet Corbin and Anselm L. Strauss. 2008. *Basics of qualitative research : techniques and procedures for developing grounded theory*. Sage, 3rd edition.

John W Ely, Jerome A Osheroff, Paul N Gorman, Mark H Ebell, M Lee Chambliss, Eric A Pifer, and P Zoe Stavri. 2000. A taxonomy of generic clinical questions: classification study. *British Medical Journal*, 321:429–432.

Barney G. Glaser and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine de Gruyter, New York.

Andreas Henrich and Volker Luedecke. 2007. Characteristics of geographic information needs. In *GIR '07: Proceedings of the 4th ACM workshop on Geographical information retrieval*, pages 1–6, New York, NY, USA. ACM.

W. R. Hersh. 2008. *Information Retrieval*. Springer. Information Retrieval for biomedical researchers.

Vahed Qazvinian and Dragomir R. Radev. 2008. Scientific paper summarization using citation summary networks. In *The 22nd International Conference on Computational Linguistics (COLING 2008)*, Machester, UK, August.

G. Salton and M. J. McGill. 1983. *Introduction to modern information retrieval*. McGraw-Hill, New York.

Karen Spark Jones. 1998. Automatic summarizing: factors and directions. In I. Mani and M. Maybury, editors, *Advances in Automatic Text Summarisation*. MIT Press, Cambridge MA.

Robert S Taylor. 1962. Process of asking questions. *American Documentation*, 13:391–396, October.

Simone Teufel and Marc Moens. 2002. Summarizing scientific articles: experiments with relevance and rhetorical status. *Computional Linguistics*, 28(4):409–445.

Elaine G. Toms. 2000. Understanding and facilitating the browsing of electronic text. *International Journal of Human-Computing Studies*, 52(3):423–452.

D Tran, C Dubay, P Gorman, and W. Hersh. 2004. Applying task analysis to describe and facilitate bioinformatics tasks. *Studies in Health Technology and Informatics*, 107107(Pt 2):818–22.

Stephen Wan and Cécile Paris. 2008. In-browser summarisation: Generating elaborative summaries biased towards the reading context. In *The 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Paper*, Columbus, Ohio, June.

Stephen Wan, Cécile Paris, and Robert Dale. 2009. Whetting the appetite of scientists: Producing summaries tailored to the citation context. In *Proceedings of the Joint Conference on Digital Libraries*.