

## An Experiment to Identify How Researchers Select Documents from Search Results

**Hagiwara, Yasuko**

Department of Library Science, Graduate School of Integrated Frontier Sciences, Kyushu University

**Wu, Meizhi**

Department of Library Science, Graduate School of Integrated Frontier Sciences, Kyushu University

**Mizutani, Emiko**

Department of Library Science, Graduate School of Integrated Frontier Sciences, Kyushu University

**Watanabe, Yukiko**

Service Division, Kyushu University Library : Head | Department of Library Science, Graduate School of Integrated Frontier Sciences, Kyushu University : Associate Professor

他

<https://hdl.handle.net/2324/1660341>

---

出版情報 : Proceedings of CiSAP Workshop. 2015, pp.8-11, 2015-12-12

バージョン :

権利関係 :

## An Experiment to Identify How Researchers Select Documents from Search Results

Yasuko Hagiwara, Meizhi Wu, Emiko Mizutani, Yukiko Watanabe, Emi Ishita,  
and Yoichi Tomiura

Department of Library Science, Graduate School of Integrated Frontier Sciences, Kyushu  
University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan  
{2FS15060R, 2FS13063W, 2FS15061N}@s.kyushu-u.ac.jp  
{watanabe.yukiko.935, ishita.emi.982}@m.kyushu-u.ac.jp  
tom@inf.kyushu-u.ac.jp

**Abstract.** Researchers must confirm the novelty of their study by investigating whether similar research was previously published. If they search documents using a high-recall search strategy, the search results will turn out to be large. This then requires a significant amount of manual reading and selection to be performed by the researcher. Therefore, a more efficient function would be preferred. This paper outlines our intended experiment to clarify the criteria for document selection and the role of sentences in abstracts, by conducting interviews and think-aloud protocols. The experiment will contribute to enhancing function to support document selection in academic databases.

**Keywords:** document selection criteria · academic database · empirical study

### 1 Introduction

Researchers typically must confirm the novelty of their study by performing a literature survey of existing research. Searching for all documents related to their research is a necessary step. However, the drawback of searching documents using a high-recall search strategy is that the number of search results will often be large and include articles of low relevance. It then becomes time consuming for researchers to select which documents meet their needs.

Ideally, a better solution would be to have a more effective function to support document selection. If academic database systems could classify search results based on researchers' document selection criteria, researchers would be able to select documents for full text view more efficiently. To support this, it is necessary to clarify how researchers select documents from search results.

This paper outlines our intended experiment. We plan to clarify the document selection criteria that researchers use when selecting documents from search results. We will do this by recording the document selection process to examine the selection criteria.

In the remainder of this paper, section 2 introduces related works for document selection criteria, while section 3 describes the proposed methods of our research.

## 2 Related Work

Some research has already been conducted in this field. Xie et al. [1] compared the evaluation of search result lists and documents, and found most participants read the abstract prior to selecting a document. Macedo-Rouet et al. [2] shed light on researchers' actual search behavior by empirically observing users during the search process through PubMed. The study found 100% of participants first read the titles, and 84% of participants then checked the abstract while selecting documents from search results. Wang et al. [3] empirically proposed the document information elements of criteria when researchers select documents. The study found the four most frequently used document information elements were title, abstract, journal, and author. The abstract was found to be the second most frequently used when evaluating the topicality and orientation/level (i.e., intellectual level and intended audience) of the documents.

The results of previous research indicate that the title and abstract are common criteria for document selection from search results. Although it is likely difficult to read all abstracts from all search results, it would be helpful if researchers were able to select documents by reading only part of each abstract. Therefore, it is important to clarify the key parts of the abstract.

Regarding the research methods used, Wang et al. [3] explained that the think-aloud protocol was the best approach to trace the thinking process. In their study of document selection decision making, they asked participants to read aloud the information in search results and to vocalize their thoughts (i.e., think aloud) while evaluating the documents. Macedo-Rouet et al. [2] also employed the think-aloud protocol in their study of article selection using the PubMed database. They asked participants to perform five identical search tasks. While searching, the participants had to describe what they did with the interface and to read aloud their queries. They were also prompted to describe what they were reading and the reasons for selecting a particular document. These existing studies enabled determination of useful information using these approaches. Therefore, we plan to conduct an empirical study by interviewing researchers as well as employing the think-aloud protocol.

## 3 Method

We aim to clarify document selection criteria by studying researchers' actual search behavior. We will focus especially on "clue sentences" within the abstract as selection criteria. We plan to ask approximately 15 engineering researchers to conduct this experiment. We will focus on engineering researchers as they reportedly spend a significant amount of time on information behavior [4]. Our method consists of three parts: 1) pre-interview, 2) experiment, and 3) follow-up interview. We intend to perform a pre-test in advance to verify the method.

### **3.1 Pre-interview (approximately 30 minutes)**

We will conduct a semi-structured interview with each of the participants to find out 1) the research situation such as research theme and research phase, 2) the information session such as information needs, criteria of document selection, intended search, and intended usage of the resulting documents. We will develop an interview sheet based on the pre-test beforehand.

### **3.2 Experiment (approximately 60–90 minutes)**

The participants will be asked to conduct an experiment using three steps. First, they will select an academic database. Second, they will perform any document search they choose. Finally, from the search results they will select documents for which they want to access the full text.

There are two conditions to the experiment. The first one is that the participants have to read at least 30 abstracts to select their chosen documents. The second one is that the maximum experiment time is 90 minutes. Participants are allowed to attempt any search within the timeframe.

Our plan is to save the search process by capturing each researcher's computer screen, and recording his or her think-aloud process. Regarding this process, researchers are asked to describe not only the criteria but also the reason for selecting a particular document. For example, the participants are expected to give a comment such as, "I'll skip article no. XX judging by the title." In this case, the criterion is the title. When it comes to abstracts, the participants will be asked to read aloud the clue sentence in an abstract as the criteria and to explain the reason as well. They would say, for example, "I'm going to read (or not to read) no. XX. I judge this is relevant based on the second sentence of the abstract."

### **3.3 Follow-up interview (approximately 30 minutes)**

We will conduct a follow-up interview to resolve and confirm any unclear comments made by participants during the experiment. We will ask them to repeat the comment or make additional explanations if necessary.

Through the above three methods, we will clarify the criteria for document selection and the role of sentences in abstracts.

**Acknowledgements.** This work is supported by JSPS KAKENHI Grant Number 15H01721.

## References

1. Xie, I., Benoit III, E.: Search Result List Evaluation versus Document Evaluation: Similarities and Differences. *Journal of Documentation*. 69(1), 49–80 (2013)
2. Macedo-Rouet, M., Rouet, J.-F., Ros, C., Vibert, N.: How Do Scientists Select Articles in the PubMed Database? An Empirical Study of Criteria and Strategies. *European Review of Applied Psychology*. 62(2), 63–72 (2012)
3. Wang, P., Soergel, D.: A Cognitive Model of Document Use during a Research Project. Study I. Document Selection. *Journal of the American Society for Information Science*. 49(2), p. 115–133 (1998)
4. Case, D.O. (ed.): *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*, 3rd ed. Emerald, Bingley, U.K. (2012)