

Proposer

Yasuko Hagiwara, Emi Ishita, Emiko Mizutani, Yukiko Watanabe, 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
TEL. +81-92-642-2780
2FS15060R@s.kyushu-u.ac.jp, ishita.emi.982@m.kyushu-u.ac.jp,
2FS15061N@s.kyushu-u.ac.jp, watanabe.yukiko.935@m.kyushu-u.ac.jp,
tom@inf.kyushu-u.ac.jp

Statement of topic

A Preliminary Study and Analysis to Identify Key Elements in Document Selection

Significance and relevance of the topic

Researchers try to obtain as many relevant documents as possible using academic databases. However, it can be very time consuming to select documents from large search results. Therefore, academic databases require more effective functions to assist with the document selection process. For instance, emphasising particular sentences in abstracts according to the information needs of researchers would aid document selection. Hence, it is necessary to clarify which parts of abstracts are used as document selection criteria. In this study, we observed search sessions in which researchers selected documents to obtain the full-text versions. Our aim was to reveal which elements are checked by researchers, and in what order. In addition, we examined the role of sentences within abstracts. In this paper, we report the results of a preliminary study involving three researchers.

Content

1 Methodology

The research plan of our study was explained in Hagiwara et al. (2015). It consisted of three steps: 1) pre-interview, 2) search session conducted by participants, and 3) follow-up interview. In the pre-interview, we asked participants their research theme, research duration, and search topic. In the search session, participants were asked to perform a document search for their search topic using particular academic databases. After the search, participants selected those documents that they wished to read in full from the search results. During the session, participants were asked to describe which elements they were checking in the search results, which sentences they used in abstracts as a clue, and their reasons for selecting particular documents. These statements were recorded, and a number of screen images were captured. In the follow-up interview, we clarified which elements and sentences in abstracts had been most useful, and the reasons for document selection.

The preliminary study involved three researchers and was conducted in November 2015 and January/February 2016. After the first pre-research, we revised the questions in the pre-interview to focus on the participants' research theme and search topic. The study was approved by the research ethics committee in the department of informatics, Graduate School of Information Science and Electrical Engineering, Kyushu University.

2 Preliminary Results

2.1 Order of elements checked for document selection

We analysed which elements were checked, and their order in the search results, to select documents. From the three participants, two patterns were identified in the order in which each element was checked. The primary element for participants A and B was the title, followed by the abstract. These elements were also identified as criteria for document selection by Wang et al. (1998) and Macedo-Rouet et al. (2012). In addition, Xie et al. (2013) reported a tendency to read abstracts before obtaining the full-text in online searches. After checking the abstracts, participant C accessed full-text versions of three documents without making a selection decision. He checked the reference in these three documents, but not the body of text. Moreover, he selected four documents from the references. According to above results, it is possible to apply our analysis to identify patterns in the order of elements.

2.2 Analysis of assigning categories to sentences in abstracts

We analysed the sentences used by participants for document selection (called “clue sentences” hereafter) and assigned them to appropriate categories. These categories were constructed according to the participants’ information needs, and have the effect of reducing the time and effort of selection. We used each researcher’s statements to analyse why they focused on particular clue sentences. Based on these reasons, we assigned each sentence to a category. For instance, participant A focused on sentences describing the method. Nevertheless, the subject field in the method caused him to discard the document. That is, he made a judgement based on the subject. If subject fields in the method were highlighted in abstracts, it would be helpful for researchers, such as participant A, to select documents from search results. However, only four categories were found in the study. We plan to examine the applicability of this approach for deriving a set of categories.

3 Future works

We plan to repeat the research with additional researchers in different fields and at different stages of research to build a more comprehensive set of categories to sentences in abstracts.

Abstract

Introduction. With regard to the selection of documents from academic database systems, we investigate the most effective search functions. For this purpose, we aim to clarify elements of the search results and the role of sentences in abstracts for document selection.

Method. Participants were asked to search for documents using academic databases. The search sessions were recorded visually (by capturing screen images) and orally by asking about the elements checked, sentences in abstracts used to select documents, and the reasons for selecting particular documents. Preliminary experiments were conducted with three researchers.

Analysis. We identified the key elements for document selection and analysed the order of elements in search sessions. We investigated the role of sentences in abstracts, and assigned categories based on the information needs and reasons for selection.

Results. Two element orders were identified. Two of the researchers used titles and abstracts as the primary elements, whereas the other researcher used titles and then accessed the full text before making a decision on selection. This third researcher also chose documents from the references in the full-text versions. Regarding the role of

sentences, the experimental data indicate there are four categories built from the information needs of the participants. Each category helps to reduce the time and effort of selection.

Conclusion. The results indicate that it is possible to apply our analysis to identify patterns in the order of elements. To clarify the role of sentences, further experiments are required to build a comprehensive set of categories.

References

- Hagiwara, Y., Wu, M., Mizutani, E., Watanabe, Y., Ishita, E., Tomiura, Y. (2015). An experiment to identify how researchers select documents from search results. *Proceedings of CiSAP Workshop, 2015*, 8-11.
- Macedo-Rouet, M., Rouet, J. F., Ros, C., Vibert, N. (2012). 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.
- Wang, P., Soergel, D. (1998). A cognitive model of document use during a research project. Study I. Document selection. *Journal of the American Society for Information Science*, 49(2), 115-133.
- Xie, I., Benoit III, E. (2013). Search result list evaluation versus document evaluation: similarities and differences. *Journal of Documentation*, 69(1), 49-80.