

Title	論文間の参照情報の抽出と利用に関する研究
Author(s)	難波, 英嗣
Citation	
Issue Date	2001-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/906
Rights	
Description	Supervisor:奥村 学, 情報科学研究科, 博士

Extraction of Citation Information and Its Applications

Hidetsugu Nanba

School of Information Science,
Japan Advanced Institute of Science and Technology

February, 2001

In a research paper, there are passages where the author of a current paper describes the essence of cited papers and the differences between the current paper and the cited papers (we call these passages “citing areas”). We call the information derived from these passages “citation information”. In this thesis, we propose a method for extracting citation information and show its applications.

In the field of citation analysis, citation relationships have been used for classifying papers, evaluating the importance of papers or journals, and analysing the relationships between research fields. However, most analyses treat all citations equally, although there are actually several reasons for citations.

In this work, we make use of citation information. With the information from citing areas, we can know the similarities and differences between the current paper and the cited papers. We can also identify the types of citation relationships that indicate the reasons for citations (we call them citation types). Citation information makes it possible to understand a stance of a paper among several related papers, or to grasp the outline of the domain. It can also contribute to the refinement of several techniques in citation analysis. We therefore attempt to extract citation information.

Extraction of citation information consists of two processes: extraction of citing areas and identification of citation types. Citing areas are defined as a succession of sentences that have a connection with the sentence that includes the citation in the paragraph. As we believed that such a connection between sentences could be indicated by some cue phrases, we used those cue phrases for citing area extraction. As a result, we obtained recall of 80 % and precision of 76 %. We then proposed a method to identify citation types automatically using several cue phrases. As a result, we obtained the accuracy of 83 %.

We use citing areas and citation types for support to write a survey article. To write a survey article, at least two processes are necessary. One is to collect papers from some domain. Another is to make clear the differences between papers. We believe that citation information is useful for both these processes. Making use of citation types, we can collect a set of papers in the same domain. Finally, we build up a system to display the citation graph of the papers. With our system, abstracts and citing areas of papers can be seen. Users of this system can easily collect papers from some specific domain and also can understand the differences between the related papers.

We also use citation types for classification of research papers. It is well known that using citation analysis makes it possible to obtain topical collections of papers. However, most previous research in citation analysis treats all citations equally. We therefore refine citation analysis by taking account of citation types. The results of our experiments showed that our method based on bibliographic coupling (“BCCT-C”) is more effective than other methods.

In this thesis, we focus on support for writing a survey article and for classification of research papers. Citation information is also generally applicable to other purposes (e.g., the analysis of research fields and the evaluation of research papers). It is also applicable to other genres of texts, such as patents and texts on the World Wide Web.

Key Words: citation analysis, extraction of citation information, multi-paper summarization, classification