論文の内容の要旨

論文題目　　A STUDY ON NAME DISAMBIGUATION USING WEB DIRECTORIES
（ウェブディレクトリを用いた人名の曖昧性解消に関する研究）

氏名　　ヴー　クァンミン

Recent advancements in computer technologies and network technologies have caused a new revolution in data communications. Cheap computers and high speed networks now can afford almost anyone to create information and to exchange information at low costs. The global internet and the World Wide Web (WWW) have brought people close together. At any second, many documents are created and updated in the WWW to reflect the dynamic changes in the real world.

As text documents in the WWW grow explosively, many users request to search for valuable information from the huge database of text documents. Among these requests of users, a certain amount of requests are to search for information related to people. Nowadays, it is very common to use search engines for the investigations and acquirements of valuable information. Upon receiving request queries from users, search engines look up in indexing databases for documents that contain query terms and return results to users. When using search engines to search for people, the issue of name ambiguity is a big problem. Since a personal name is often shared by several people, results from search engines often contain documents relevant to several people. Therefore, users have to manually look for the person of interests in the result sets.

In our research, we targeted the name ambiguity problem in web searches and developed a new method to separate the person of users' interests from other people automatically. We proposed to use web directories as additional information to improve the disambiguation performance. Web directories were collections of documents categorized in some topics. Several well-known directories already existed in the WWW and we could use them at low preparation costs. We proposed two approaches to utilize information in web directories. In the first approach, we used the directories themselves to extract contexts relevant to people in search result documents and used these contexts to disambiguate people in result documents. In the second approach, we first preprocessed web directories to extract topics contained in web directories. Then, we used these extracted topics for the extraction of contexts related to people in documents containing ambiguous names. We used several well-known web directories to disambiguate ambiguous names in real web documents. The experiment results showed that our approaches extracted contexts of people effectively and performed better than other name disambiguation approaches that used the vector space model method and the named entity recognition method.