

論文内容の要旨

博士論文題目

Studies on Improving Retrieval Accuracy in Web Information Retrieval
(Web 情報検索における検索精度改善に関する研究)

氏 名

Kazunari Sugiyama (杉山 一成)

(論文内容の要旨)

The rapid spread of Internet made it possible for anyone to access a variety of information on the World Wide Web (WWW) easily. Web search engines are now one of the most popular methods for finding valuable information. However, users are not satisfied with the retrieval accuracy of search engines. The following two facts cause this problem.

- (1) Accurate indexing of Web pages is not performed,
- (2) Most search engines cannot provide search results that satisfy each user's information need respectively.

The aim of this thesis is to solve those problems.

- (1) We propose approaches to refine the TF-IDF scheme for a target Web page by the contents of its hyperlinked neighboring pages. Experimental results show that, generally, more accurate feature vectors of a target Web page can be generated by utilizing the contents of its neighboring pages at levels up to second in the backward direction from the target page.
- (2) Typical Web search engines return the same result to all the users who submitted the same query. Since each user has his own information needs, the search result should be adapted to each user respectively. We propose approaches to achieve it. Experimental results show that better search systems that adapt to a user's preferences can be achieved by constructing user profiles based on modified collaborative filtering.

The proposed approaches contribute for indexing a target Web page more accurately, and for allowing each user to perform fine-grained search that satisfy his/her information need.

氏 名	杉山 一成
-----	-------

(論文審査結果の要旨)

2003 年 12 月 26 日に開催した公聴会の結果を参考に、2004 年 2 月 19 日に本博士論文の審査を行った。次のとおり、本博士論文は、提案者が独立した研究者として、研究活動を続けていくための十分な素養を備えていることを示すものと認める。

杉山一成は、本博士論文において、Web ページに書かれている内容的な視点と、検索を行なう利用者側の視点から、Web 情報検索における検索精度の改善手法を提案した。本論文の成果は次のとおりである。

(1) Web ページの内容をより正確に表現するために、ハイパーリンクで結ばれた隣接ページの内容も用いた。これは、通常の文書の特徴付ける方法としてよく使われる TF-IDF 法の改良である。実験の結果、提案手法は、TF-IDF 法と比較して、最適な場合で 5%弱の適合率の改善が確認された。

(2) 検索エンジンの利用者の検索要求は、利用者ごとにそれぞれ異なることに着目し、各利用者の長期的な嗜好と短期的な嗜好を利用者の閲覧履歴から利用者に負担をかけることなく自動的にとらえ、このようにして得られた嗜好に適応させて検索結果を提示する手法を提案した。実験の結果、通常の検索エンジンが出力する結果と比較して、15%程度の適合率の改善が確認された。

本研究は、Web 情報検索における検索精度の改善手法の提案であり、独創性が高く、しかも実用的で、研究開発現場において高い貢献があると評価する。よって、本論文は博士（工学）の学位論文として価値あるものと認める。