A Method of Re-ranking Web Search Results Using their Hidden Hyperlink Structure

Nara Institute of Science and Technology
Kazunari Sugiyama
Background

Transition of Search Engines

- The first generation
  - Based on term written in Web pages
- The second generation
  - Based on hyperlink structure of Web pages

Problems

- A large number of Web pages are returned.
- Web pages which are not relevant to user’s query are often ranked at upper position.

It is required to describe the feature of a Web page considering contents among Web pages connected by hyperlink structure.
System overview

WWW

Search Engine

Search Results
1. http://www.abc.co.jp
3. http://www.u-ja.ac.jp
4. http://www.c-de.co.jp

User queries

Re-ranking Search Results
1. http://www.u-ja.ac.jp
2. http://www.c-de.co.jp
3. http://www.abc.co.jp

Generate Query Vector, $Q$

Calculate similarity between $Q$ and $W'$

Generate modified Feature Vectors, $W'$

Link Analysis of each Web page
Link analysis for generating feature vector (in Web space)

Making K clusters from hidden Web Pages

[Hidden Web pages]
Generating feature vector (in vector space)

Vector of searched Web page

Modified vector of searched Web page

Clusters Generated from hidden Web pages

Modifying vector of searched Web page using centroid vector of cluster generated from hidden Web pages
Experimental results

Compare retrieval accuracy
• existing search engine
• conventional tf-idf method
• proposed method

L and K means the number of hierarchy from searched Web page and clusters, respectively.
Conclusion and Future Work

Conclusion

A method for re-ranking Web search results using their hidden hyperlink structure is proposed.

The contents of a Web page is summarized within 2 or 3 links away from the searched Web page.

Future Work

Using both backward and forward links.

There are many types of Web pages. Therefore, the feature vector of a Web page needs to be made depending on each link type of the Web page.