

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

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

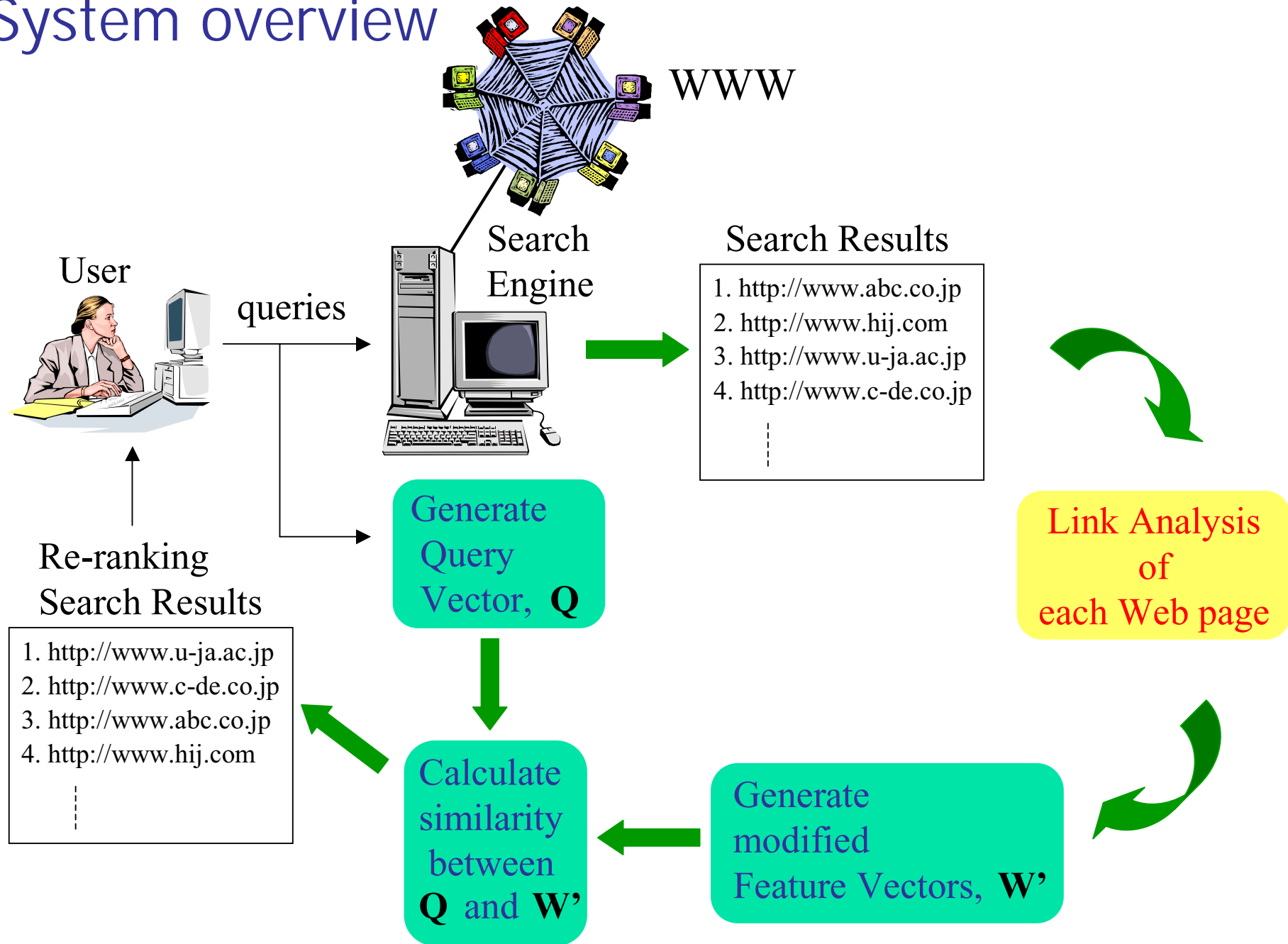
Problems

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

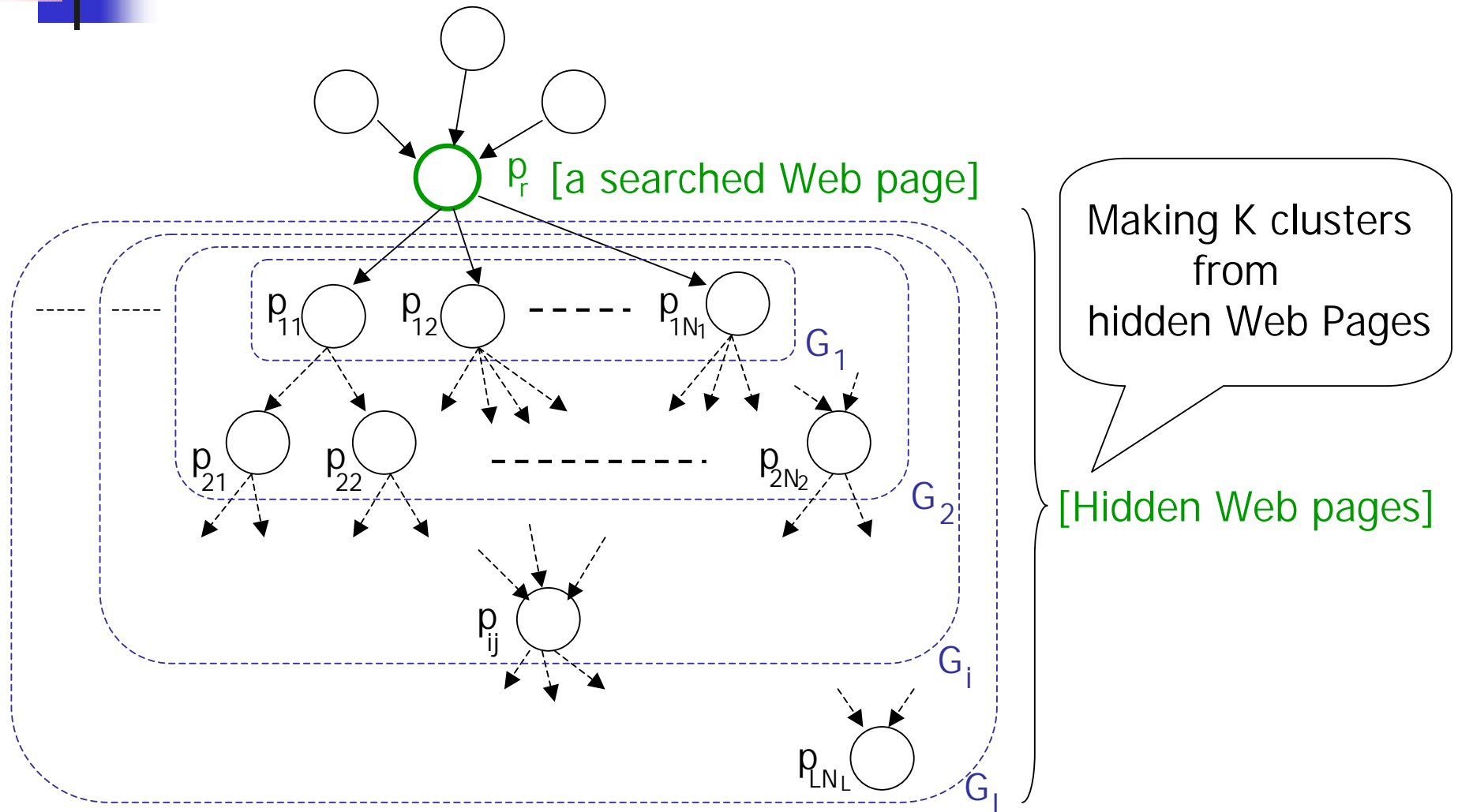


It is required to describe the feature of a Web page considering contents among Web pages connected by hyperlink structure.

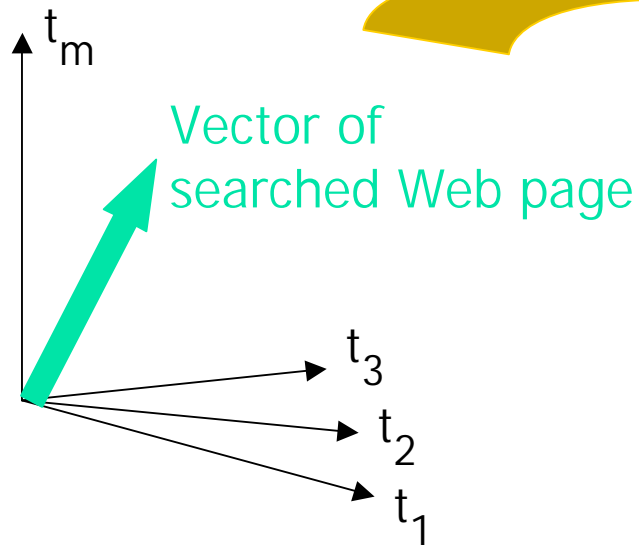
System overview



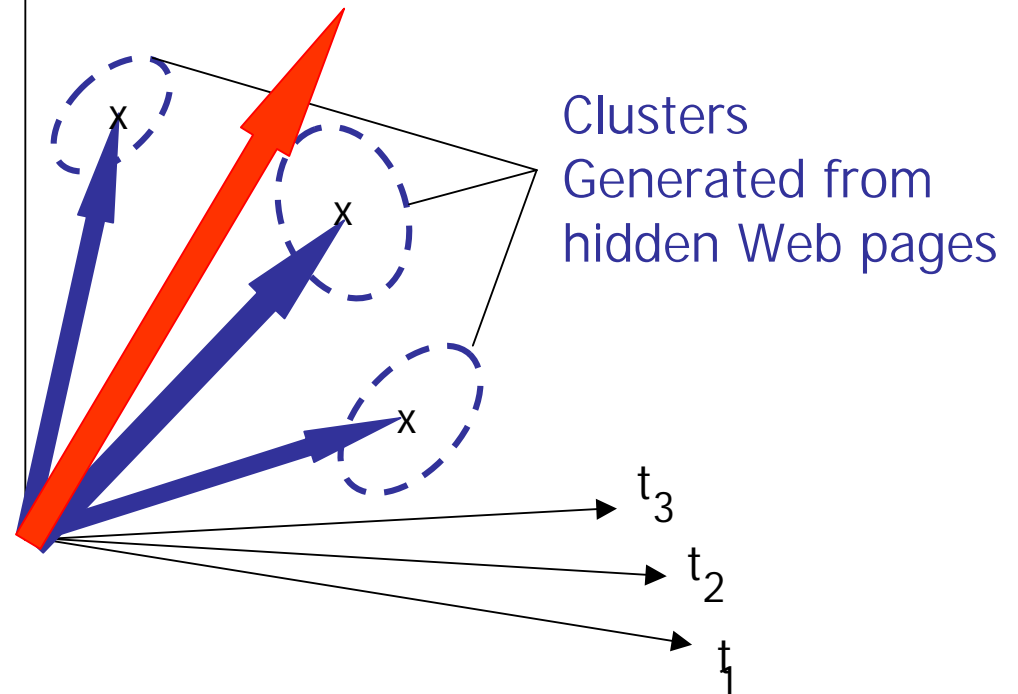
Link analysis for generating feature vector (in Web space)



Generating feature vector (in vector space)



Modified vector of searched Web page



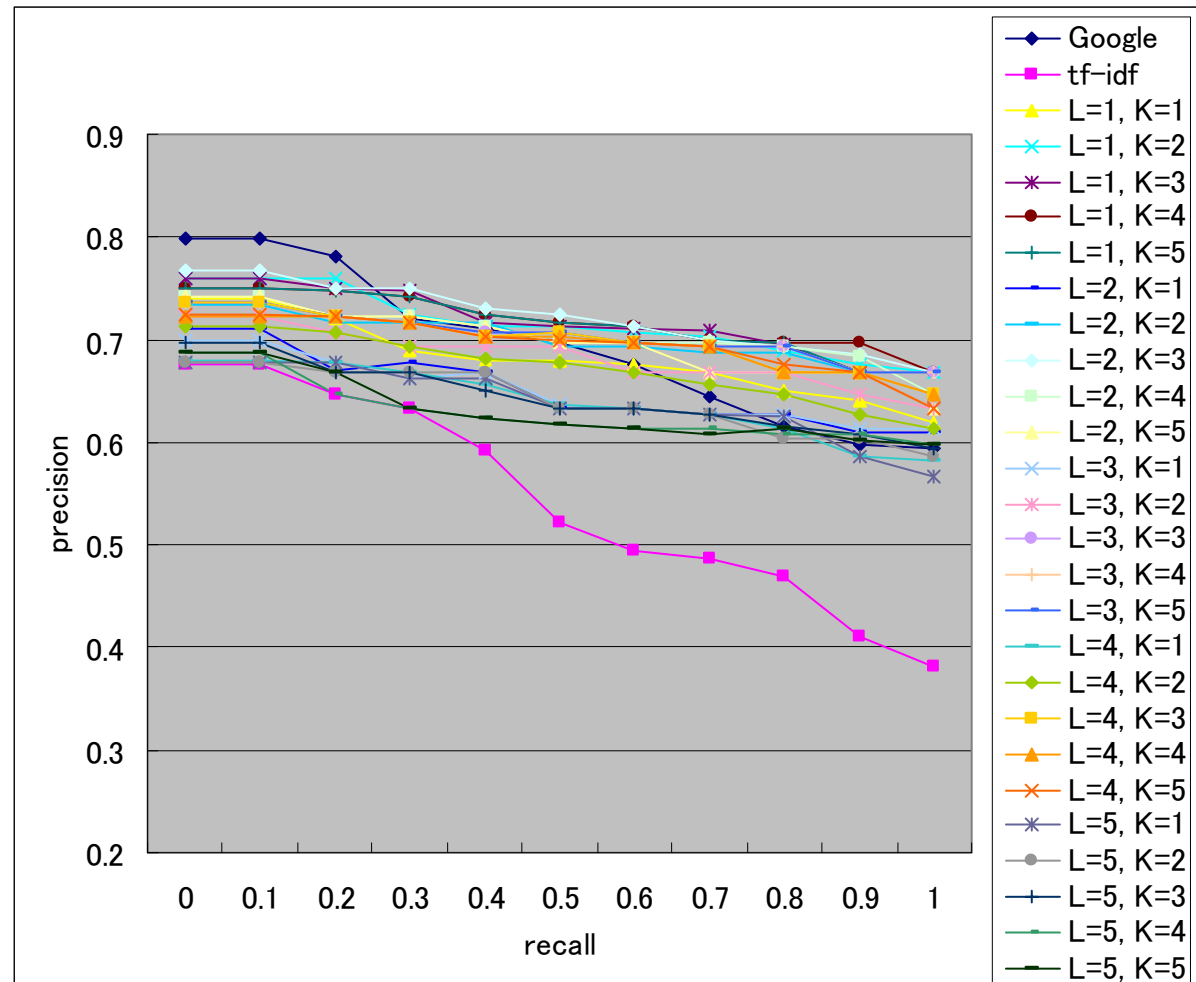
Modifying vector of searched Web page using centroid vector of cluster generated from hidden Web pages

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

n Conclusion

- n A method for re-ranking Web search results using their hidden hyperlink structure is proposed.
- n The contents of a Web page is summarized within 2 or 3 links away from the searched Web page.

n Future Work

- n Using both backward and forward links.
- n 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.