Interactive Web search by graphical query refinement

Junji Tomita
NTT Cyber Space Laboratories
1-1 Hikari-no-oka, Yokosuka-Shi
Kanagawa, 239-0847
Japan
tomita@isl.ntt.co.jp

Genichiro Kikui
NTT Cyber Space Laboratories
1-1 Hikari-no-oka, Yokosuka-Shi
Kanagawa, 239-0847
Japan
kikui@isl.ntt.co.jp

ABSTRACT
We propose a new Web search system that helps users clarify their information needs through interaction. The system represents the user's information needs using a query graph: a set of graph objects. This graph is presented to the user and interactively refined until the user is satisfied with the documents retrieved. In this way, the user can find the desired documents even if his original information needs were unclear. This article describes the concept and outlines the system.

Keywords
Web search, information needs, graphical interface, interaction, relevance feedback.

1. INTRODUCTION
People often try to search the Web with poorly specified information needs. Unfortunately, most Web search systems, including those with powerful ranking algorithms, do not satisfy these people. This is because existing search systems expect users to give well-specified queries, which is difficult for the people whose information needs are vague. Relevance feedback [1], which modifies the original query based on the user's judgment of previous search results, may support the users somewhat. Simple relevance feedback, however, prevents the user from understanding what the relationships between query form and documents are judged to be relevant, because it automatically modifies the original queries while concealing the modification process from the user. A *penetrable interface* that allows the user to control the relevance feedback search process has been shown to be effective [2].

In addition to controlling the search process, an interface that effectively represents user's information needs and clarifies the needs through interaction is indispensable in supporting users with vague information needs.

We propose a new search system with a human interface that uses graphical forms to visualize the user's information needs and navigates the user to the desired information by manipulating the forms.

2. Representing Information Needs
Boolean queries and natural language queries are widely used in the field of Information Retrieval (IR). Boolean queries are a powerful way to represent clear information needs, but they are not suitable for representing vague information needs. Natural language queries appear more suitable for this purpose, but accurate evaluation of natural language queries is still difficult.

In the field of Artificial Intelligence, Semantic Networks, which are graphs of linking concepts, have been used to represent ideas. It seems intuitive for humans to think in terms of using the relationships of concepts. Accordingly, we decided to utilize a graph of linking concepts to represent information needs. Each graph, called a *query graph*, sets a word at each node; the relation between words is represented as a link. A link can be thought of as Boolean 'and' or 'proximity'. Two independent nodes that have no link indicate Boolean 'or'. For example, Figure-1 depicts the information needs of 'travel guide to Asia or Japan, or transportation by train'. In this way, query graphs can visualize the information needs.

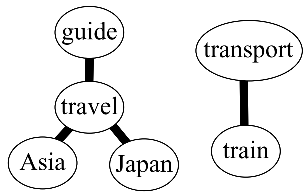

Figure-1: A Query Graph.

3. Interactive Query Graph Refinement
The user can interactively clarify his/her vague information needs, by looking at and editing the query graph. Figure-2 depicts the user interface of our search system. The steps in refining a query graph are as follows.

(1) The user inputs sentences (keywords or documents) representing his/her information needs. The system displays the initial query graph made from the inputs.

(2) The user edits the query graph to match his information needs by removing and/or adding nodes and/or links.

(3) The system measures the relevance score of each document against the modified query graph. The score is defined using the similarity between the query graph and the *subject graph* of each document.

This is called *subject graph matching* [3]. In a subject graph, each node represents a word in the document, while a link represents a relation between nodes. Details of the matching algorithm and experimental results are shown in [3][4]. Subject graph matching can measure the scores more precisely than the conventionally utilized vector space model [5], because the matching algorithm incorporates term relations calculating the scores.

(4) The system ranks search results in descending score order, and displays their titles to the user interface with relevant checkboxes beside the titles that allow the user to select documents as desired.

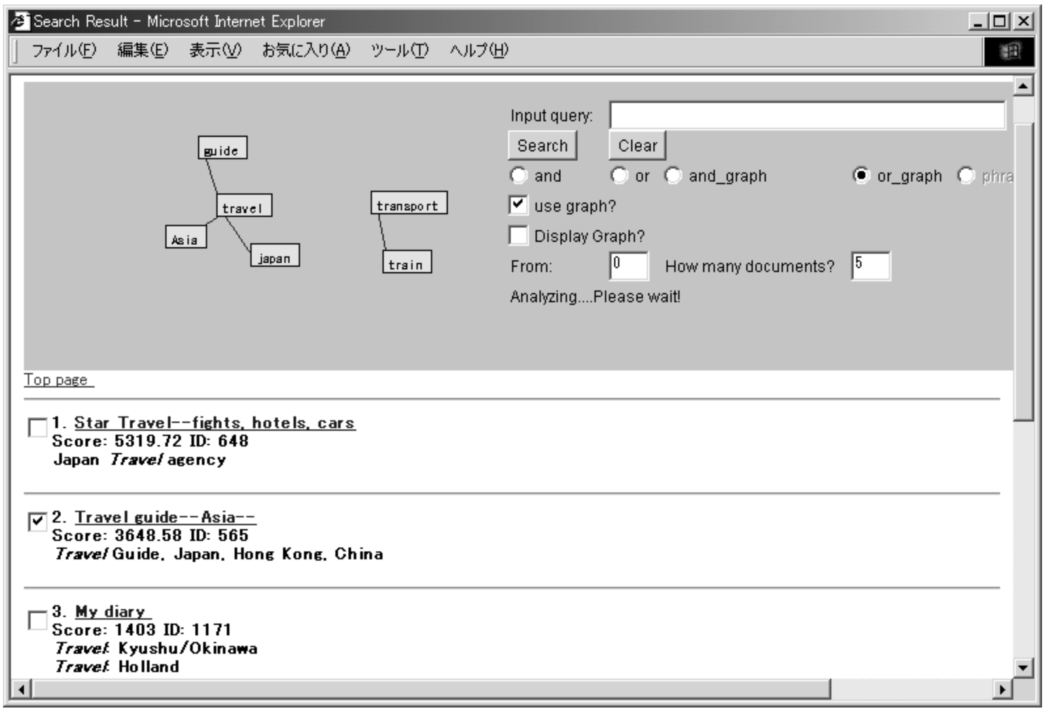

Figure-2: User interface of our system.

(5) The user selects a checkbox if the document appears to be relevant to his/her information needs.

(6) 1. The system makes a subject graph of the entire content of each selected document. 2. It then extracts a *neighbor subgraph*, from each subject graph, including nodes matched to the query graph and their neighboring nodes. 3. It merges these subgraphs and the query graph. Figure-3 depicts these steps.

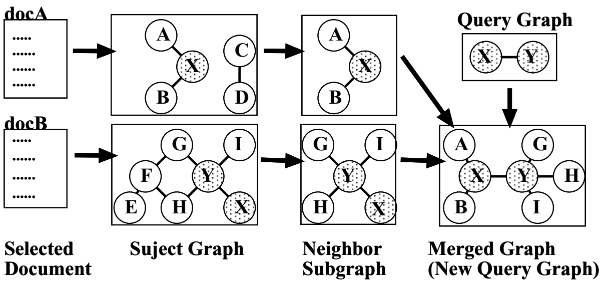

Figure-3: The steps of making a new query graph.

(7) The system displays the merged graph, which can then be used to initiate a new query.

(8)Steps (2)～(7) are repeated interactively until the user is satisfied with the search results.

In this way, the user clarifies his/her vague information needs by selecting relevant documents, and editing query graphs interactively. The user obtains the desired search results in response to a query graph that reflects the *clarified information needs* of the user.

4. Disscusion

As a preliminary evaluation, we showed a query graph and a list of terms, both representing the same information needs, to 5 users. All users indicated that the query graph was more intuitive and made the needs easier to understand than the list of terms. Comprehensive evaluations are planned of the interactive search process as well as the query graph.

The proposed query graph has weights on its nodes and links, but these values are currently not visible, nor are they editable. Size or color clues could be utilized for visualizing them. The current query graphs use one word to represent one concept, but it is reasonable to expect that some concepts will actually need several words. We want to enhance the query graphs to handle such concepts, and different types of relations such as 'agent', 'object' and 'place'.

5. Conclusion

We presented a new search system that can clarify vague information needs. Because its process is visible and controllable, users can efficiently find the desired documents. Subject graph matching can measure the score between a query graph and each document precisely.

6. REFERENCES

[1] William B.Frakes and Ricardo Baeza-Yates. *Information Retrieval, Data Structures & Algorithms*, Prentice Hall, 1992

[2] J. Koenemann and N. J. Belkin. A Case For Interaction: A Study of Interactive Information Retrieval Behavior and Effectiveness, *CHI'96, 205-212*, 1996

[3] Junji Tomita and Yoshihiko Hayashi. Improving Effectiveness and Efficiency of Web Search by Graph-based Text Representation, *Poster Proceedings of WWW9*: 54-55, 2000.

[4] Junji Tomita and Hiroshi Takeno. Proposal for a new IR system using subject graph and word's weighting by the relation (in Japanese), *IPSJ-SIGFI*, 52(3): 17-24, 1998.

[5] G. Salton and A. Wong and C. S. Yang. A Vector Space Model for Automatic Indexing, *Communications of the ACM*, 18:613-620,1975.