Interactive Document Search with *DualNAVI*

Yoshiki Niwa, Makoto Iwayama, Toru Hisamitsu, Shingo Nishioka, Akihiko Takano, Hirofumi Sakurai, and Osamu Imaichi

Central Research Laboratory, Hitachi, Ltd.
Hatoyama, Saitama 350-0395, Japan

{yniwa,iwayama,hisamitu,nis,takano,hirofumi,imaichi}@harl.hitachi.co.jp

Abstract

DualNAVI is a user-interface for information retrieval which provides users with rich interaction methods based on two kinds of duality, *dual view* and *dual query types*. The dual view interface provides summary information of retrieved results in the form of topic word graph and enables feedback by selected words. The dual query types supports search by key documents and enables feedback by selected documents. We evaluated the effectiveness of these two types of feedback by using NACSIS IR test collection. The results were significantly positive for both types of interaction.

1 Introduction

For document retrieval systems, a variety of assisting functions have been designed and developed so that users might readily reach a desired document set as soon as possible. Guidance and feedback are primary functions among them.

DualNAVI (Nishioka et al.1997) supports both of assisting functions in a integrated manner. First, *DualNAVI* provides summary information of retrieved results in the form of topic word graph and enables feedback by words which meet user's interest. It also supports search by key documents as well as search by keywords, and this enables feedback by documents selected by users.

In order to evaluated the effectiveness of these two types of feedback, we submitted following three sets of search results for the *ad hoc IR task*.

- (S) simple use of original query sentence.
- (A) feedback by selecting related articles.
- (W) feedback by deletion of worthless words and addition of extra keywords from topic word graph.

In actual retrieval processes, these two types of feedbacks are used in a variety of combination depending on the nature of query and actual retrieved results. So the above evaluation method may be over simplified. But we think the above evaluation will be helpful for us to consider about the effect of interactive methods.

In the following part of this paper, we first introduce the interface model of *DualNAVI* in Sect. 2, and describe the experimental methods in Sect. 3, and the evaluation results in Sect. 4.

2 *DualNAVI* interaction model

As to the guidance, conventionally, a method has generally been executed, comprising proposing information relevant to input search quearies. For example, a method is illustrated, comprising storing a database representing relations between words, such as thesaurus, and retrieving form the data base, a set of words relevant to the input keywords. A method is also suggested, comprising automatically generating a data of relevant words using co-occurrence statistics (Schatz, Johnson, and Cochrane1996).

(Fowler and W.Dearholt1990) also proposed a method comprising displaying a search word and words relevant to the word in a network structure, on the basis of the co-occurrence statistics data between words.

Another method is proposed, comprising automatically extracting relevant information from search results and providing the information to a user. For example, Scatter / Gather method (Hearst and Pedersen1996), comprises automatically classifying a retrieved document group (clustering) and displaying the topic words per each class.

However, these conventional methods are still not enough for meeting the large variety of helpless situations which we encounter in the real searching process.

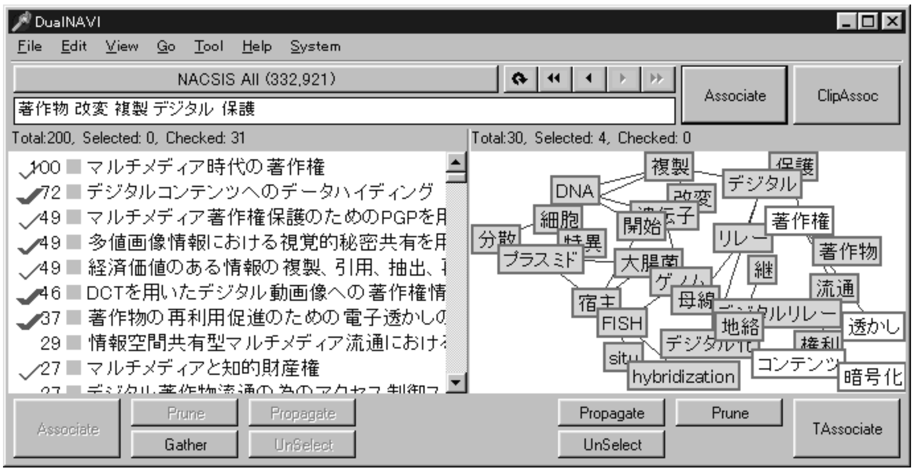

Fig. 1. Navigation Interface of *DualNAVI* (Nishioka et al.1997)

2.1 Duality for Intensive Interaction

The advantageous features of *DualNAVI* interaction model is summarized by two kinds of duality: one is *dual view* of retrieved results and the other is *dual query types.*

First, *dual view* is composed of two types of presentation of retrieved results. One is the ordinary title list, and the other is the topic word graph (see figure 1). A topic word graph is composed of words characterizing the retrieved results and works as a summary of the set of retrieved documents.

In general, *duality* means two different types of things working complementarily, that is, each component complements what the partner does not have. In this case, we can get individual information from the title list and get more abstract or summarized information from the topic word graph. So this pair is a typical example of duality.

On the other hand, *Dual query types* means search by key documents together with search by keywords. When our search demand is clear and can be expressed by some specific keywords, keyword search works effectively. But it is not always the case. When it is difficult to get proper keywords, search by key documents (search by the similarity of documents, or associative search) is helpful. So these are also complementary partners.

Dual view bridges Dual query types

The dual view and dual query types are not just two isolated features, but they are closely related. Dual query types can work effectively only with dual view framework. Figure 2 illustrates how they relates.

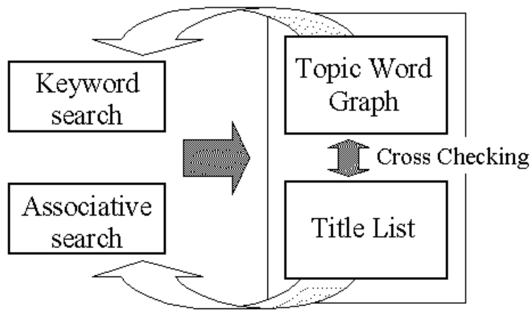

Fig. 2. Dual view bridges dual query types

We can start with either of keyword search or associative search. then we get the dual view of retrieved results (the right side). If we find interesting articles in the title list, we can proceed to next associative search with the found articles as key documents. If we find interesting words in the topic word graph, we can proceed to next keyword search by using the found topic words.

DualNAVI also supports cross-checking function between title list and the topic word graph, which enables users to easily check what topic words appear in selected articles and what articles contain selected topic words.

Technical details concerning the associative search and the generation of topic word graph will be briefly mentioned as parts of experimental methods in the following section.

3 Experimental Methods

As we mentioned in the introduction, we made three sets of search results and examined their performance. One is the simple automatic case where we used the query sentences as they are. Other two sets of results are those derived by using two major feedback methods supported by *DualNAVI* . One is the results of feedback by selecting relevant articles, and the other is the results of feedback by selecting relevant topic words.

3.1 Simple automatic method (S)

The system first segment the query sentences into word sequence and only content words are used for query. The search is basically done as the disjunction of the words ($w_1 \vee ... \vee w_n$) and then the score of each retrieved document is calculated by the sum of usual $tf \cdot idf$ term weight over the keywords.

Segmentation of query sentences are done by morphological analyzer ANIMA (Sakurai and Hisamitsu1999) which was also used for making the index file.

3.2 Feedback by related articles (A)

Feedback procedure by selected articles is an ordinary process of relevance feedback. Users can select each related article by clicking the small squre region just left of the title. After selecting related articles, The *Associate* button at the left bottom corner starts the associative search with the selected articles as key documents. The details of associative search is as follows.

Associative Search

At the first stage of associative search, a designated number of characteristic words are selected from the retrieved set of documents, based on the following scoring of each word w.

$$p(w|q) \ / \ p(w)$$

Here q means the set of key documents, and $p(w|q)$ is the probability of appearance of word w in the key documents. The number of selected words, in this experiment, was set to 200. The set of selected words for key document set q is written by $K(q)$.

Next retrieval process is basically the same as the keyword search with the 200 selected words $K(q)$. The relevance score of each retrieved document d is given by

$$R(d,q) = \frac{1}{log(length(d))} \cdot \sum_{w \in K(q)} tf(w|d) \cdot \frac{p(w|q)}{p(w)}$$

3.3 Feedback by topic words (W)

If we find some words in the topic word graph which meet the current interest of search, then we can use these words in the next step query. The third set of results is for examining the effect of this feedback process.

figure 1 is the case for query id 0068, “Digital copyright books”. After removing worthless words in the query sentences such as 方法 (method), 述べる (state), and 論文 (paper), current key words are 著作物 (copyright book), 改変 (change), 複製(reproduction), デジタル (digital), and 保護 (protection).

In the word graph, we find some related words such as 著作権 (copyright), 透かし (watermark) or, 暗号化 (encryption). So we added these related topic words to the query words and did the next search. Because we judged the new results were better than the previous one and there are no additional words to be added in the new topic word graph, we determined the results as the result of this feedback method.

Generation of Topic Word Graph

The process of topic word graph generation is composed of three stages. The first stage is the extraction of topic words based on the word frequency analysis over the retrieved set of documents. Next stage is the process of generating links between extracted topic words based on cooccurrence analysis. The last process is assigning xy-coordinates for each topic word on the display area.

The selection of topic word is based on the following score of each word w:

$$\frac{df(w) \text{ in the retrieved documents}}{df(w) \text{ in the whole database}}.$$

In general, it is difficult to take the balance of high frequency words and low frequency words by using single score. In order to over come this difficulty, we adopted frequency-class method, in which, all candidate words are first roughly classified by their frequencies, and then proper number of topic words are taken from each frequency class. As for the details, see (Niwa et al.1997).

In the link generation stage, each topic word X is linked to another topic word Y which has larger document frequency in the retrieved set and maximizes the co-occurrence index $df(X\&Y)/df(Y)$. In the last graph drawing process, the y-coordinate is given according to the document frequency of each word (in the retrieved set). The x-coordinate has actually no meaning, but only assigned in a way preventing overlapping. Since the vertical direction reflects the frequency of each word, the graph can work as the hierarch of topics in the retrieved set.

WAM based computation

In *DualNAVI* , all types of major retrieval computations are done with compressed matrix type data structure, which we call WAM (word-article-matrix). WAM is composed of pair of matrix, word to article matrix and article to word matrix. At each place of a matrix is stored the frequency of each word in each article. Word to article matrix works as the index file when keyword search is done. Article to word martix is used when topic words are chosen. The number of different words (used for word index) from 330,000 NACSIS abstracts was about 180,000. The size of WAM data was about 350 megabytes.

In order to count the word frequencies for making WAM, the original text data was segmented into words by ANIMA (referred above). The text was segmented into 60 million words in about 140 minutes on Compaq Alpha 300MHz.

4 Evaluation

We evaluated the three set of search results using the model answers of relevance presented by NACSIS collection.

First we scored each result by two types of scoring indices; one is the maximum average of precision and recall, which we here call global index and the other is a kind of index evaluating the quality of top ranking items. The details are as follows:

Global index: In this experiment, we retrieved 1000 items for each query. The global index is for measuring the total performance over these 1000 items. For each top n items, we calculate a kind of average (or average of inverse) of precision and recall by the following formula:

$$\frac{2 * precision_n * recall_n}{(precision_n + recall_n)}$$

Here the suffix $_n$ means the value is for top n items. We used both A and B rated articles as correct answers. Then we take the maximum of above value over n changing from 1 to 1000.

Top-ranking index is for measuring the quality of retrieval results with highest scores. We used such index because the top-ranking quality is most sensitive to the users' actual intuitive evaluation. Here we defined the top ranking as top 20th, and counted the number of A-rated and B-rated articles in top 20th. Then the index is given by

$$2 * (\text{number of A}) + (\text{number of B}),$$

and then normalized by dividing this value by the ideal-value. (When there are more than 20 A's in the answer set, then the ideal-value is 40. If there are 10 A's and 20 B's, then the ideal case is 10 A's and 10 B's, so the ideal value is 30.)

Figure 3 shows the results. The left side numbers show the query IDs, the left part of the graph shows the global index and the right part shows the top-ranking index.

In both cases, the simple baseline method is shown by ||, and the feedback-by-article is shown by ∘, and the feedback-by-word is shown •.

~~~~~~~~~~~~~

We found that there is a very strong correlation between the scoring by the global index and by the top ranking index. So these figures are reliable enough so that we can consider the performance of each search methods based on these data.

What is clear from this results is that both feedback methods had siginificantly positive effect over the simple case.

Among 53 queries, the feedback-by-articles method had remarkably positive effect for 11 cases, and had negative effect only for 4 cases. As for the feedback-by-words method there were 14 cases of remarkably positive effect and 3 cases of negative effect.

However one serious mistake of our experiments was that we failed in separating the effect of adding extra key words from the topic word graph and the effect of removing worthless words.

So the pure effect of feedback-by-words is not clear from these results and therefore we can not compare the effect of two feedback methods.

One thing we can say here about the relation between two feedback methods is that these methods seem to have independent effects. Because the eleven cases for which the feedback-by-articles had big effect and the seven cases for which the feedback-by-words had big effect shares only 3 cases.

In the following part, we will consider the results of feedback-by-words in detail, especially concerning the effect of deleting worthless words and the effect of adding extra topic words. We think this analysis will be useful in the future work for more pure evaluation of the effect of topic-word presentation.
~~~~~~~~~~~~~

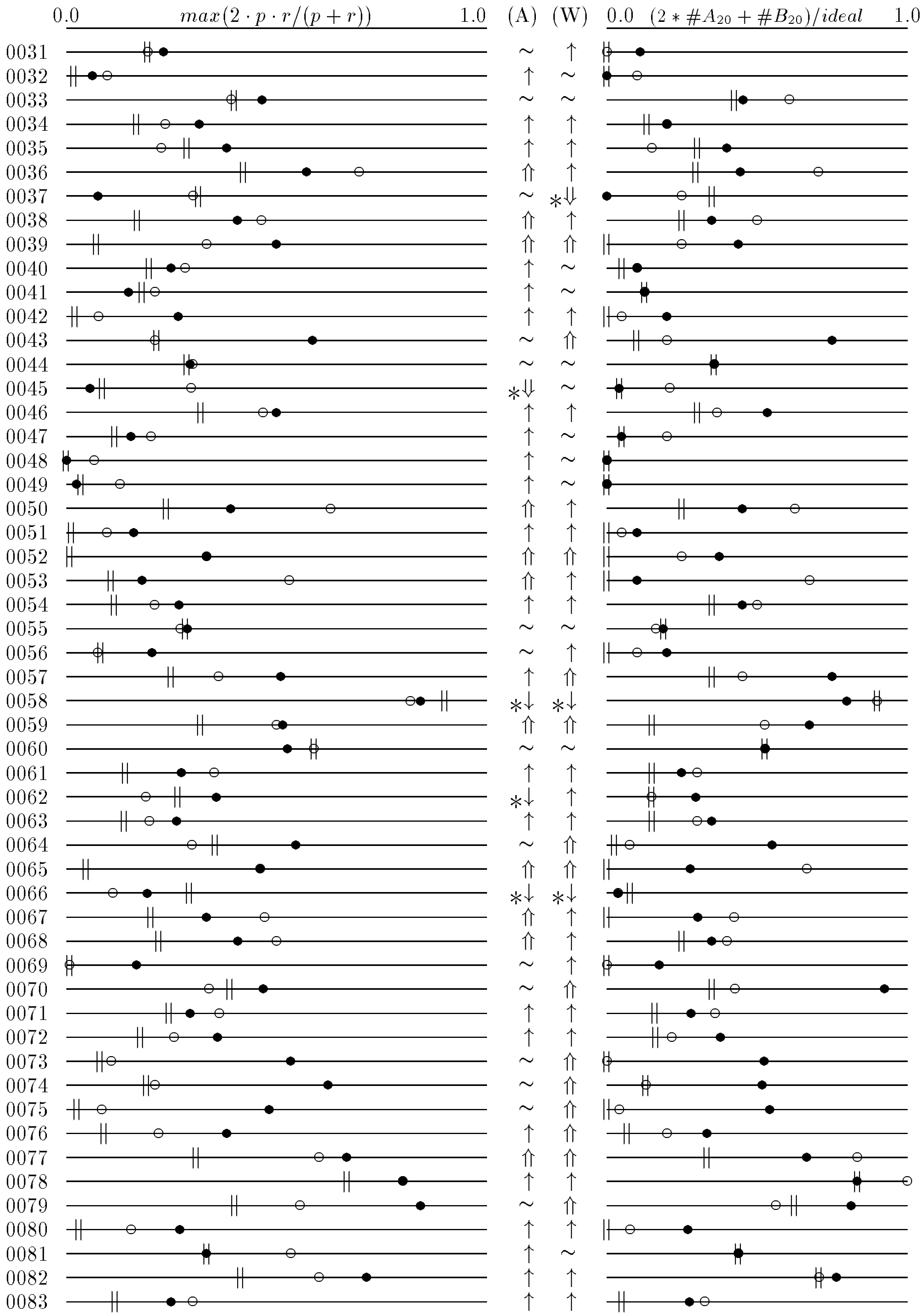

Fig. 3. Global score (left) and Top-rank score (right)

||: (S) simple, ∘: (A) feedback by articles, •: (W) feedback by words

The central two columns (A) and (W) show the effects of feedbacks.

Table 1. keywords added and deleted in the feedback-by-words method

QID	words from original query	Effect	added from topic words	non-trivial deleted words
0031	マルチキャスト 品質 複数 通信 データ 制御	↑	QOS QoS	
0032	ミドルウェア コラボレーション 構築 ネットワーク	~		
0033	メディア 同期 ネットワーク 分散 環境	~		
0034	TCP 無線 通信 適用 制御	↑		
0035	図書館 電子 分散 環境	↑	検索 ネットワーク	
0036	モバイル グループウェア 問題点 環境	↑		
0037	バッファ 転送 有効 データ 利用	*⇓	ルータ TCP IP スループット ATM ネットワーク	
0038	TCP IP スループット ATM 網 論文 通信 特性	↑		
0039	WWW トラヒック 分析 モデル 化	⇑	トラヒックモデル	
0040	精細 論文 生成 画像 高	~	解像度 画素	
0041	精細 医療 超 応用 技術 画像 高	~	SHD	
0042	動 次元 通信 画像 処理	↑		
0043	イメージセンサ 知能 圧縮 動 画像 化	⇑		
0044	透かし 埋め込む 電子 成分 周波数 画像	~	DCT ウェーブレット	
0045	リング アクセス 多重 ネットワーク 制御 型	~		
0046	ATMR リセット 改良 周期 アルゴリズム	↑	スロッテドリング LAN	
0047	マルチキャスト 信 実際に 配 遅延 性能 時間 データ 評価 特性	~		
0048	ホーム エリア ネットワーク 実現 手法	~		
0049	トポロジー 有効 ネットワーク 情報 性	~		
0050	将棋 知能 人工 応用	↑		
0051	インフラストラクチャ プロジェクト インターネット 世代 次	↑		
0052	ソーラーエネルギー 自動車	⇑		
0053	発する 人体 電波 無線 機器 検証 及ぼす 影響	↑	曝露 FDTD 電話 携帯 頭部 電磁	
0054	ギガビット ファイバ 通信 光	↑	光通信 高速	
0055	LaTeX 全文 SGML 索引 文書 階層 論理 データベース 検索 記述 論文 効率 高速 技術 利用 構造 的	~		
0056	部門 異 共有 情報	↑	協調	設計 製造 保全 人工 ライフサイクル 知識
0057	アイデア 創造 思考 支援 過程 生成 設計 モデル	⇑	発想	
0058	Zipf 法則 応用	*↓	ジップ	
0059	シソーラス 維持 構築 自動	⇑		
0060	占領 事情 教育 問題	~		
0061	語 資源 階層 言語 自動 抽出 関係	↑	シソーラス 辞書	
0062	ボランティア 生涯 学習	↑		
0063	コンピュータ 文字 言語 多	↑	コード	国語
0064	生涯 学校 活用 事例 施設 学習	⇑		
0065	遺伝 検索 応用 アルゴリズム 画像	⇑	GA	
0066	図書館 大学 探索 行動 建築 空間 利用	*↓	館内 閲覧 施設	視点
0067	避難 学校 災害 実態 自然 施設	↑	被災	
0068	著作物 改変 複製 デジタル 保護	↑	透かし 著作権 コンテンツ 暗号化	無断 使用
0069	教科 小学校 補助 コンピュータ 利用 的	↑		
0070	高齢者 行動 調査 特性	⇑	外出 日常	
0071	使用料 著作物 課金 著作権 体系 電子	↑		
0072	博物館 公開 資料 電子 システム	↑		
0073	精子 マウス 形成 解析	⇑	精巣	分子 生物学
0074	G1 酵母 因子 周期 遺伝子 制御	⇑		細胞
0075	組立 機械 運動 作業 シミュレーション	⇑		製品 物体
0076	多面体 干渉 検出	⇑		
0077	点字 数式 翻訳 処理	⇑	点訳	
0078	テロメラーゼ 不死 テロメア 哺乳 老化 動物 関与 活性 細胞	↑		
0079	アミロイドタンパク アルツハイマー β 蓄積	⇑	アミロイド アルツハイマー病	
0080	動物 損傷 メカニズム 再生 神経 分子	↑		脊椎
0081	性分化 性差 哺乳類 メカニズム 脳 レベル 分子	~		
0082	マラリア 抗 活性	↑		薬剤
0083	骨 動物 メカニズム レベル 分子 形成	↑	吸収	脊椎

Table 2. Effect of additional keywords

additional keywords	total	⇑	↑	[⇑ & ↑]	∼	↓	⇓	[∼ & ↓ & ⇓]
yes	24	7 (29%)	11 (46%)	18 (75%)	3 (13%)	2	1	6 (25%)
no	29	7 (24%)	13 (45%)	20 (69%)	9 (31%)	0	0	9 (31%)
	53	14 (26%)	24 (45%)	38 (72%)	12 (23%)	2	1	15 (28%)

4.1 Feedback by words

Table 1 shows what words were added and what words were removed from the original query sentences in the process of feedback by words. The second column shows the words which were in the original queries and not removed in the interaction. The fourth column shows the words which are taken as additional keywords from the topic word graph after deleting worthless words in the segmented query.

The fifth column shows words deleted from the original query. Here we only listed non-trivial words such as "molecular biology" deleted from query 0073. Besides these non-trivial words, we also deleted clearly worthless words such as 的 (-like), 論文 (paper), 欲しい (want) if they appeared in the segmented word sequence of a query.

The central column (column 3) shows the effect of the change of keywords. Except 3 contrary cases (qid = 0037, 0058, 0069), and some almost equivalent cases (marked with ∼). The effect was mostly positive (38 among 58 cases), and for 14 cases, the effect was considerably large (they are marked with thick upward arrows). But as we mentioned earlier, we need more investigation for evaluating the effect of topic word graph in the interaction.

First, what we should remark here is that it is just 24 cases among 58 queries that we added extra key-words from the topic word graph. As for the rest 34 cases, only worthless or harmful words in the original queries were removed. The comparison of effects for the former added cases and latter non-added cases are shown in table 2.

The percentage of positive effect (⇑ *or* ↑) is slightly higher for the cases extra keywords were added. This result may be natural because we add keywords only when we find good keywords in the topic word graphs.

4.1.1 Contrary effect of added keywrods

A problematic phenomenon for added cases is that there are three negative effect cases, whereas there was none for non-added cases. Let us see individual cases.

qid = 0037 (⇓) バッファ制御 (buffer control)

The original query is " データ転送におけるバッファの有効利用について述べた論文はないか。" (effective use of buffer in data transmission).

Added keywords are ルータ (router), TCP, IP, スループット (throughput), ATM, and ネットワーク (network).

These seem to be related to the topic, but the first three words (router, TCP, and IP) had a bad influence of making the results out of focus from the main topic, *buffer*. This is because these three words are too general words in this text database.

qid = 0058 (↓) Zipfの法則の応用 (application of Zipf's law)

In this case, Japanese expression ジップ of Zipf appeared in the topic word graph, and was added to the keywords. But the result was slightly lower than the simple case, although the difference was almost negligible. This is because there is one article containing ジップ in the sense of Zip disk. Other appearances of ジップ are all accompanied by Zipf. So the addition of ジップ had only negative effect.

qid = 0066 (↓) 大学図書館の建築空間 (building space of university library)

The original query is " 利用者の探索行動という視点から大学図書館の建築空間を扱った研究。(building space (?) of university libraries from the view point of users' behavior for searching)"

Added keywords are "館内 (inside of building), 閲覧 (reading), 施設 (institution)".

These words were added without very clear reason. Because some of the high ranking articles were concerning digital libraries, we chose these words as words associated with real libraries with an expectation that these words would reduce the number of digital library related articles. But it was not the case. In this case, the negative effect seems mainly caused by the un-clear understanding of the intention of the query.

4.1.2 Deletion of non-trivial words

There were several cases where the improvement through interaction was caused by the deletion of non-trivial words in the original query sentence. They are as follows (See also table 1):

qid = 0073 マウスにおける精子形成を，＜分子生物学＞的手法を用いて解析した文献

qid = 0074 G1期における＜細胞＞周期制御にはたらく因子，遺伝子で，酵母から取得されたものについて述べた文献

qid = 0080 ＜脊椎＞動物において神経再生の際に、損傷を受けた神経に起きる分子メカニズムを解析した文献。

The deleted non-trivial words are bracketed by < ... >. These words, <molecular biology> in 0073, <cell> in 0074, and <spinal column> in 0080 are not at all worthless words and highly related to the main topic of each query. Whether they work positively or negatively seems to be determined by a delicate combination of several factors and seems very difficult to predict automatically.

Here we would like to make a comment that *DualNAVI* interface was useful when we decided what words sould be removed. In case of qid = 0073, the initial word graph showed that the most dominant topic was *molecular biology. Mouse* and *sperm* were also major topics but they appeared in separate branches of the graph. So we could judge that *molecular biology* was working as a bridge connecting a group of articles only related to *mouse* and almost disjoint group of articles only related to *sperm.*

After removing *molecular biology* from the query words, *mouse* and *sperm* appeared in the graph with direct link between them. That means they shares contexts in a large part of retrieved articles. We could also know from the title list that the results improved much.

5 Conclusion

By using the NACSIS IR test collection, we examined the effect of two types of feedback method supported by *DualNAVI* interface model. One is the feedback by selecting related articles and the other is the feedback by selecting related topic words. By comparing these results with that of the simple non-interactive method, we could make sure that both interaction methods have positive effect for almost all queries.

Among 53 queries, the feedback-by-articles method had remarkably positive effect for 11 cases, and had negative effect only for 4 cases.

The feedback-by-words method was applied to 24 cases and had remarkably positive effect for 7 cases, but had negative effect for 3 cases.
As for the rest 29 cases, the interaction was done only by removing clearly worthless words and some additional words which seemed have bad influences. In these cases, there were 7 cases of remarkable improvement and no cases had negative effect.

One serious mistake of these experiments was that we failed in separating the effect of adding extra key words from the topic word graph and the effect of removing worthless words. So the pure effect of feedback-by-words is not clear from these results and therefore we can not compare the effect of two feedback methods.

In the future, we would like to evaluate the pure effect of using topic word graph and also investiage the effect of combined use of feedback by articles and feedback by words, That is, how the feedback by articles can improve the results after feedback by words, and vice versa.

References

Fowler, Richard H. and Donald W.Dearholt, 1990. *Information Retrieval Using Pathfinder Networks*, chapter 12, pages 165–178. Ablex.

Hearst, Marti A. and Jan O. Pedersen. 1996. Reexamining the cluster hypothesis: Scatter/gather on retrieval resuts. In *Proceedings of SIGIR'96*, pages 76–84.

Nishioka, Shingo, Yoshiki Niwa, Makoto Iwayama, and Akihiko Takano. 1997. *DualNAVI*: An information retrieval interface. In *Proceedings of WISS'97*, pages 43–48. (in Japanese).

Niwa, Yoshiki, Shingo Nishioka, Makoto Iwayama, and Akihiko Takano. 1997. Topic graph generation for query navigation: Use of frequency classes for topic extraction. In *Proceedings of NLPRS'97*, pages 95–100.

Sakurai, Hirofumi and Toru Hisamitsu. 1999. A data structure for fast lookup of grammatically connectable word pairs in japanese morphological analysis. In *Proceedings of ICCPOL'99*, pages 467–471.

Schatz, Bruce R., Eric H. Johnson, and Pauline A. Cochrane. 1996. Interactive term suggestion for users of digital libraries: Using subject thesauri and co-occurrence lists for information retrieval. In *Proceedings of ACM DL'96*, pages 126–133, Maryland.