

# Fisheye Matching: viewpoint-sensitive feature generation based on concept structure

Y. Takama<sup>a,\*</sup>, M. Ishizuka<sup>1,b</sup>

<sup>a</sup>Department of Computational Intelligence & System Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8502, Japan

<sup>b</sup>Department of Information & Communication Engineering, School of Engineering, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

Received 7 April 1999; received in revised form 20 March 2000; accepted 23 March 2000

## Abstract

Recent rapid growth of information environment such as the Internet makes it easy for us to get vast information. On the other hand, “information overflow” is becoming a serious problem. To cope with such a problem, we have extended the normal Vector Space Model (VSM) to reflect the users’ viewpoints more clearly. We call this new matching method the Fisheye Matching method, which generates the features related to the users’ viewpoints based on the concept structure of an electronic dictionary. In the Fisheye Matching method, the users’ viewpoints are expressed as a set of word groups, each of which corresponds to a certain concept in the concept structure. Each concept in the dictionary has heading information, and the users can grasp their viewpoints easily from such information. Experimental results on information retrieval show that the Fisheye Matching method can not only retrieve documents in which the users take interest, but also supply them with useful information on their viewpoints. © 2000 Elsevier Science B.V. All rights reserved.

**Keywords:** Document ordering; Concept structure; Vector Space Model

## 1. Introduction

Recent rapid growth of information environment such as the Internet makes it easy for us to get vast information. On the other hand, “information overflow” is becoming a serious problem. That is, available information resources are too much for a user to utilize, and sometimes confuse his/her activity. For example, the more documents (papers, news articles, etc.) a user has related to his interest, the harder it is for him to read each document carefully; in other words, the harder it is to grasp the concept of each document and to find out various relations between documents. Consequently, computers are expected to play an important role as an intelligent amplifier of human’s information-processing ability in the near future.

We assert that ordering documents while reading is an effective way of dealing with the vast collection of documents, and the systems that assist such processes should be able to find relations among documents based on the users’ viewpoints/interests.

From this point of view, we have suggested a new vector generation/matching method called *Fisheye Matching* [1], which extends the existing Vector Space Model (VSM) to use the features generated dynamically based on the users’ viewpoints/interests.

In this paper, we present an aspect of the Fisheye Matching method and show its ability by applying it to information retrieval. The Fisheye Matching method can cope with the users’ variable judgement on the similarity among documents, which is useful for both retrieving ability and their viewpoint information presented to them in a readable manner. This paper is organized as follows. We first explain the Fisheye Matching method compared to the existing VSM. In Section 3, we describe how to find semantic groups as features of a *Fisheye vector* from a concept dictionary. Experimental results are shown in Section 4, and in Section 5 we discuss the characteristics of our research by comparing them with other researches.

## 2. Fisheye Matching

### 2.1. What is document ordering

When a user gets some ideas from papers or articles, he

\* Corresponding author. Tel.: + 81-45-924-5682; fax: + 81-45-924-5676.

E-mail addresses: takama@hrt.dis.titech.ac.jp (Y. Takama), ishizuka@miv.t.u-tokyo.ac.jp (M. Ishizuka).

<sup>1</sup> Tel.: + 81-3-5841-6347; fax: + 81-3-5841-8570.

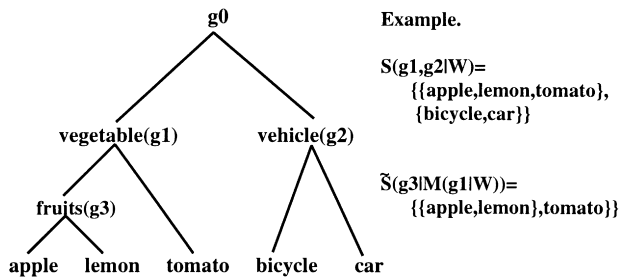


Fig. 1. An example of concept structure.

may organize his thoughts by relating incoming information with knowledge that already exists in his mind. This process is getting harder for him in proportion to the volume of information that he considers, and it is useful to illustrate the concept structure on a paper or on a display, which leads to reducing his confusion.

Employing external representations of a concept structure has been adopted by a lot of creative thinking methodologies and computer-assisted tools [2–4], which are expected to have the following effects:

- A user can be conscious of his own viewpoints or problems that have been ambiguous before [3].
- A user can notice the points of similarity or difference clearly by comparing his concept structure with others [4].
- Modifying an illustrated concept structure makes it easier to get an idea by trial and error [5].

The effects listed above are also important when we design systems that help the users order a large number of documents. Namely, the concept space formed by documents, which have been already read, is useful for readers not only to grasp his current interest/viewpoint, but also to decide which document to read next.

The VSM [6,7] has been used mostly in the field of document retrieval because of its simplicity and retrieving capability. However, the VSM seems to have the following defects in handling the users' viewpoints/interests. First, the VSM treats the users' interests almost like a black box, because the users' interests are reflected as the weight of features (words in documents are generally used) in the feature vector (often called *profile*), from which it is hard for the users to find out clearly what they are interested in. Second, each feature of the VSM is assumed to be orthogonal to each other, but this assumption is not necessarily satisfied because there are several kinds of semantic relation among words such as synonyms and antonyms. For example, when we are interested in some vehicle, there should be some relations between the document which includes the word 'car' and the document which includes the word 'bicycle'. However, we cannot find such a relation on the normal feature vector space.

Then we suggest a new document matching method, the Fisheye Matching method, which extends the VSM to cope

with the problem discussed above. The Fisheye Matching method generates features that reflect a user's current interests dynamically based on the concept structure extracted from an electronic dictionary. The details will be discussed in the rest of this paper.

## 2.2. Definition of Fisheye Matching

As discussed above, the Fisheye Matching method is a kind of extended VSM, which generates features by considering semantic relations among words. Namely, instead of adopting each word as an independent feature, the vector space used in the Fisheye Matching method is generated by combining words, which seem to have the same meanings from the user's viewpoint, into one feature, or by employing only words that relate to the user's interest as independent features. This vector space can be reconstructed dynamically whenever the user wants to change his viewpoint. The vector of each document in this vector space is called the Fisheye vector, which is calculated from the basic feature vector used in the normal VSM.

We assume that a user's viewpoint consists of topics he has in mind while reading documents. Usually, one document belongs to several topics, and the users can read a document from several different viewpoints. Because some topics are subsumed by the other topics, a document can be viewed more specifically/generally even if the user has the same viewpoint.

We utilize the concept structure of the EDR electronic dictionary developed by Japan Electronic Dictionary Research Institute, Ltd.<sup>2</sup> to extract topics. We assume that all topics have the corresponding concepts in this dictionary, and the concepts are used as the basic units of generating features of the Fisheye vector.

Fig. 1 shows an example of a concept structure that exists among 5 words, 'apple', 'lemon', 'tomato', 'bicycle' and 'car'. In Fig. 1, 'apple' and 'lemon' are included in the concept named 'fruits', and these two words and 'tomato' are included in the concept named 'vegetable'. 'Bicycle' and 'car' are included in the concept named 'vehicle', and all 5 words are included in the most general concept in this concept structure ( $g_0$  in Fig. 1).

In the Fisheye Matching method, concepts in the concept structure are treated as sets of words, which are used as the basic units to select words or to generate features based on the users' viewpoints.

If the user wants to grasp documents only roughly to classify them into documents about vegetables and about vehicles, his viewpoint may consist of topics about 'vegetable' and 'vehicle'. In such a case, the concepts 'vegetable' and 'vehicle' should be used as features of the Fisheye vector instead of using these 5 words as independent features. Namely, the words 'apple', 'lemon' and 'tomato' correspond to the same feature 'vegetable', and words

<sup>2</sup> <http://www.ijjnet.or.jp/edr>.

‘bicycle’ and ‘car’ are considered to be the same feature ‘vehicle’. We call an operation that combines words included in the same feature based on the concept structure a *Shrink* operation.

On the contrary, if the user is interested in the topic about vegetables or fruits and he wants to discriminate fruits from vegetables, the words ‘apple’ and ‘lemon’ should be considered to be part of the same concept ‘fruits’, and the concept ‘fruits’ and the word ‘tomato’ should be treated as independent features in the Fisheye vector. In this case, the words ‘bicycle’ and ‘car’ have no relations with the user’s viewpoint, and they should not be treated as features.

To select only words that relate with the user’s viewpoint corresponds to *Magnify* only an interesting part of the concept structure. Accordingly, Fisheye vectors are generated by performing these two operations on the basic feature vectors.

The sets of feature in the Fisheye vector generated by the Shrink operation and the Magnify operation are defined as follows.

$$S(g_1, \dots, g_m|W) = \{f_i|f_i = \{w_j|w_j \in g_i \cap W\}, 1 \leq i \leq m\}, \quad (1)$$

$$M(g_1, \dots, g_m|W) = \{w_i|w_i \in \cup_{j=1}^m g_j \cap W\}, \quad (2)$$

$$\tilde{S}(g_1, \dots, g_m|W) = S(g_1, \dots, g_m|W) \cup \bar{M}(g_1, \dots, g_m|W), \quad (3)$$

$$\bar{M}(g_1, \dots, g_m|W) = \{w_i|w_i \in W - \cup_{j=1}^m g_j\}. \quad (4)$$

Here,  $W = \{w_1, w_2, \dots, w_n\}$  indicates the set of words and the  $g_i$ s are called semantic groups, which correspond to concepts in the concept structure discussed above. These semantic groups are considered to be the background knowledge when generating the Fisheye vector. The right-hand side of Fig. 1 shows an example of some operations, where  $W = \{\text{apple, lemon, tomato, bicycle, car}\}$ .

$S(g_1, \dots, g_m|W)$  and  $M(g_1, \dots, g_m|W)$  are the feature sets generated by the Shrink and the Magnify operations, respectively. These two operations are the essential operations of the Fisheye Matching method, which can examine the similarity among documents from different viewpoints; for example, if the similarity of the two documents based on the Shrink operation is high, while that based on the Magnify operation is low, it can be concluded that those two documents have the same topic, but are not similar in the category of the topic.

In this paper, the Fisheye Matching method is applied to information (document) retrieval. However, the Shrink and the Magnify operations cannot be applied to information retrieval as they are, because of the decrease of features used as the dimension of the vector space.  $\tilde{S}(g_1, \dots, g_m|W)$  is defined as a Shrink operation extended by the Magnify operation (notice that the tilde has no mathematical meaning here.). From the viewpoint of retrieving ability, the Shrink operation plays a role in generating features reflecting the

users’ viewpoints, while the Magnify operation complements the number of features.

The Fisheye vector of a document  $d$ ,  $F_d(v_{f_1}, \dots, v_{f_m})$ , is calculated from the basic feature vector  $O_d(v_1, \dots, v_n)$  based on the features generated by above operations. When generating the Fisheye vector, the value  $v_{f_i}$  for the corresponding feature  $f_i$  of the semantic group  $g_i$  is the sum of the values  $v_j$  for all corresponding features of the words included in  $g_i$  as,

$$v_{f_i} = \sum_{w_j \in g_i} v_j. \quad (5)$$

The Fisheye Matching method can evaluate the relations among documents in the same way as the normal VSM [6,7]. For example, the value  $\text{Sim}(d_i, d_j)$  indicates the similarity between documents  $d_i$  and  $d_j$ , which is the cosine between them. Here,  $F_i(v_{i_1}, \dots, v_{i_m})$  and  $F_j(v_{j_1}, \dots, v_{j_m})$  are the Fisheye vectors of document  $d_i$  and  $d_j$ , respectively.

$$\text{Mag}(d_i) = \sqrt{\sum_{j=1}^m v_{ij}^2}, \quad (6)$$

$$\text{Sim}(d_i, d_j) = \frac{\sum_{k=1}^m v_{ik} \cdot v_{jk}}{\text{Mag}(d_i)\text{Mag}(d_j)} \quad (7)$$

### 3. Finding semantic groups

#### 3.1. Concept structure of electronic dictionary

We use the EDR electronic dictionary to calculate semantic groups. As for the concepts, the EDR dictionary prepares three dictionaries: a concept structure dictionary that defines the hierarchical relations among concepts; a concept description dictionary that defines the other relations among concepts; and a concept headings dictionary that explains each concept. The concept structure dictionary is used to calculate semantic groups. All semantic groups specified as arguments of the operations to generate the Fisheye vectors have the corresponding concepts in the EDR concept structure dictionary. In our experiment, concepts including 2 to 256 words are used for the semantic groups. In case there exists a word that does not belong to any group, it is handled as a special semantic group to which only one word belongs.

The concept structure or the thesaurus of the electronic dictionary is considered to be what reflects our common sense, which is suited for the background knowledge to generate the features related to the users’ viewpoints.

#### 3.2. Semantic group finding algorithm

Although the users can specify the set of semantic groups manually themselves, the users’ viewpoints usually consist of several topics (=semantic groups) and it is a time-consuming job for them to specify all semantic groups

that reflect their viewpoints. In this section, we show how to find the set of semantic groups that represents a user's viewpoint automatically.

Here,  $D_P$  indicates the set of documents which the user considered to be interesting from his viewpoints,  $D_N$  denotes the set of documents which the user considers not to be interesting. Also,  $v_{ji}$  denotes the corresponding value to the  $word_i$  in the basic feature vector  $O_{dj}$  of the document  $d_j$ . Initially,  $Wlist$  is set to include all words which construct the basic feature vector space.

1. Calculate the weights  $w_i$  for each  $word_i$  in the  $Wlist$  by using the equation below. Notice that this equation is similar to that used in relevance feedback of the normal VSM [7].<sup>3</sup>

$$w_i = \alpha \frac{1}{|D_P|} \sum_{d_j \in D_P} v_{ji} - \frac{1}{|D_N|} \sum_{d_j \in D_N} v_{ji} \quad (8)$$

2. Select the word  $word_k$ , which has the largest positive value in  $Wlist$ . Quit if such a word does not exist.
3. Calculate the set of semantic groups,  $G_k = \{g_i | (word_k \in g_i) \wedge (\forall word_j \in g_i, word_j \in Wlist) \wedge (\forall word_j \in g_i, w_j \geq 0)\}$ .
4. Calculate the weights  $W_{g_i}$  for each group  $g_i$  in  $G_k$  by using the equation below. Go to step (6) if  $G_k = \emptyset$ .

$$W_{g_i} = \frac{1}{|g_i|} \sum_{word_j \in g_i} w_j \quad (9)$$

5. Select the group  $g_l$ , which has the largest value  $W_{g_l}$  in  $G_k$ , and let  $Wlist = Wlist - \{word_i | word_i \in g_l\}$ . Go to the step (2).
6. Let  $Wlist = Wlist - \{word_k\}$  and go to step (2).

Our algorithm described here has been designed by extending the Relevance Feedback algorithm [7], which is one of the most famous methods in the VSM literature to reflect the users' viewpoints. The Relevance Feedback algorithm calculates the users' profiles based on the assumption that the important words have large weights calculated by Eq. (8).

Extending this assumption, we assume that the concept which represents the users' interests includes words that are important for the users, and have large weights  $w_i$  (calculated by Eq. (8)).

It is noticed when a certain set of semantic groups (which corresponds to the users' viewpoints) is specified as the argument of the Magnify and the Shrink operation described in Section 2.2, the Magnify operation can view documents in detail from only the specified viewpoints. On the

contrary, the Shrink operation can grasp the documents roughly as for the specified viewpoints.

#### 4. Experimental results

In this section, we illustrate the experimental results of using the Fisheye Matching method for document retrieval.

The task is to retrieve the news articles relating to medical topics from online news articles of various topics. We collected 218 online news articles from the WWW in which 100 articles were related to medical topics.<sup>4</sup>

From all collected articles, we extract 1588 words that are not included in a stop-word list and have a noun entry in the EDR English word dictionary.<sup>5</sup> It is noticed that verbs, adjectives, and so on are also important as features. The reason why only nouns are used in this paper is that the concepts in the word dictionary for words other than nouns are thought not to be suitable for the users' viewpoints.

We used these extracted words as the basic features. Based on these words, we obtained 655 semantic groups from the EDR concept structure dictionary as described in Section 3.1.

The experiment was carried out by the following steps. In each experiment, all the following steps were performed within 5 s on a workstation Sun Ultra2 (320 MB memory).

1. Input several news articles as positive and negative samples.
2. Find the set of semantic groups  $\{g_1, \dots, g_n\}$  from these samples based on the algorithms described in Section 3.2.
3. Generate the Fisheye vector of each news article by performing the  $\tilde{S}(g_1, \dots, g_n | W)$  operation.
4. Generate the profile  $p$  by using the value calculated by Eq. (9) as the weight of each group feature, and the value calculated by Eq. (8) as the weight of each word feature.
5. Calculate  $\text{Sim}(p, d_i)$ , where  $d_i$  indicate articles not used as positive or negative samples in step (1).

The aim of the experiment is to evaluate the abilities of the Fisheye Matching method when applied to information retrieval.

Information retrieval is adopted as the subject, because the retrieving ability can be examined independent of other factors, and common metrics such as precision are defined. Furthermore, information retrieval is also important as a preprocessing for document ordering.

The advantages of the Fisheye Matching method, when applying to information retrieval, are as follows.

1. Matching among different words in the same concept can be done; this cannot be done essentially by the VSM.
2. The users' viewpoints can be handled explicitly, and it is

<sup>3</sup>  $\alpha$  is a constant to define the balance of the effect of positive samples and that of negative samples, which is usually determined empirically. The value of 1.0 is used in the experiments of Section 4.

<sup>4</sup> Whether a news article relates to medical topic or not was judged from the categorization at the news-site.

<sup>5</sup> All words were stemmed in advance.

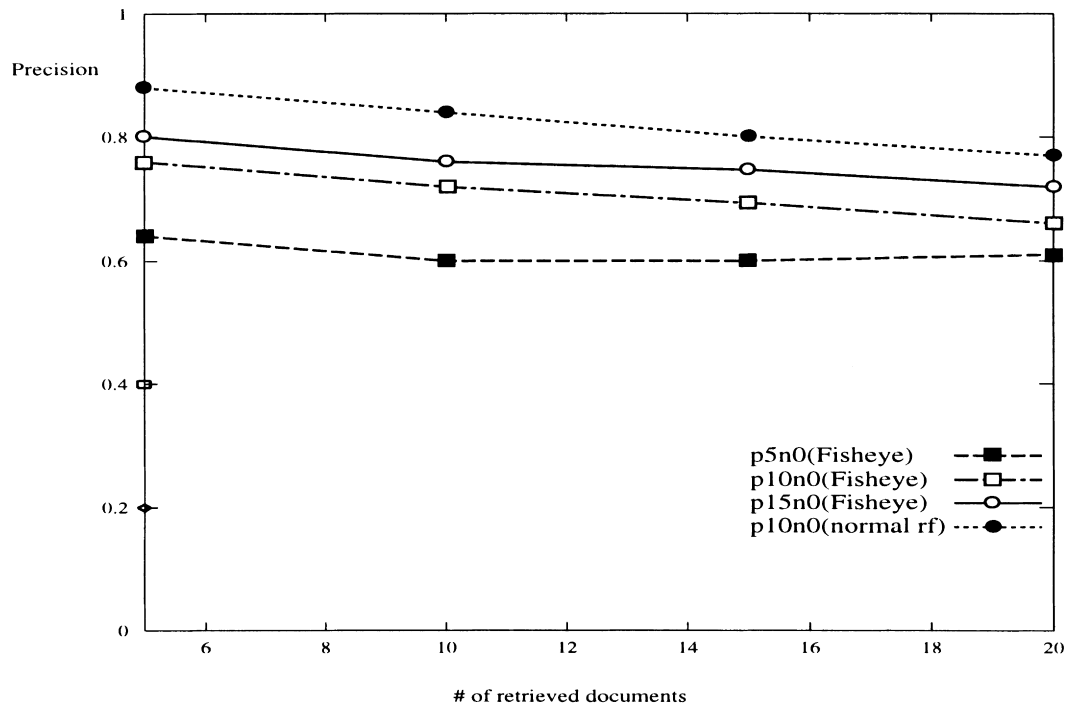


Fig. 2. Comparison of Fisheye Matching method (Fisheye) against the normal Relevance Feedback method (normal rf) when only positive samples are inputted.

easier for the users to modify their profiles by the semantic groups than to modify the weight of each word, as forced by the VSM.

The former advantage is brought by the Shrink operation, and it is noticed that the retrieving ability may not be always improved. In the rest of this section, we examine if the Fisheye Matching method can achieve at least the same retrieving ability as that of the existing VSM with the Relevance Feedback method. The  $\tilde{S}(g_1, \dots, g_m|W)$  operation, an extended version of the Shrink operation for information retrieval, is used in the experiments described here.

As for the latter advantages, the characteristics of the semantic groups extracted as viewpoints, such as the number and size and their heading information, are examined.

Table 1

Average number of extracted semantic groups and words per groups through experiments

Condition	No. of groups	No. of words/groups
p5n0	23.4	9.2
p10n0	32.8	8
p15n0	38.4	7.2
p10n5	26.2	6.8
p10n10	23.2	5.9
p10n0 (Japanese)	42.6	9.3
p10n10 (Japanese)	29.6	5.8

#### 4.1. Examine retrieving ability

As for condition (1), we used a value of precision  $P$  [8], which is calculated by the following equation:

$$P = \frac{\text{No. of correct docs in retrieved docs}}{\text{No. of retrieved docs}} \quad (10)$$

Fig. 2 shows the precision when only positive samples were given as input. The number of positive samples varied from 5 to 15 articles. Five trials are done for each experimental condition (the number of positive/negative samples), and the lines in this figure indicate the average score of each condition. The same set of positive and negative samples is given for all methods with the same condition. The vertical axis shows the precision value, and the horizontal axis shows the number of retrieved documents.

In Fig. 2, the title 'pxny(method)' indicates that  $x$  positive samples and  $y$  negative samples were given to *method*. As for the *method*, 'Fisheye' indicates the method of this paper, and 'normal rf' indicates the Relevance Feedback method usually used in the normal VSM.

In Fig. 2, we can see that the precision value of the Fisheye method gets close to that of the Relevance Feedback method according to the number of given positive samples. As the characteristics of extracted and used semantic groups, the average number of extracted groups and words per extracted semantic group are shown in Table 1. It is confirmed from this table that when more positive samples are given, more semantic groups are extracted and used. On the contrary, the average number of words belonging to each

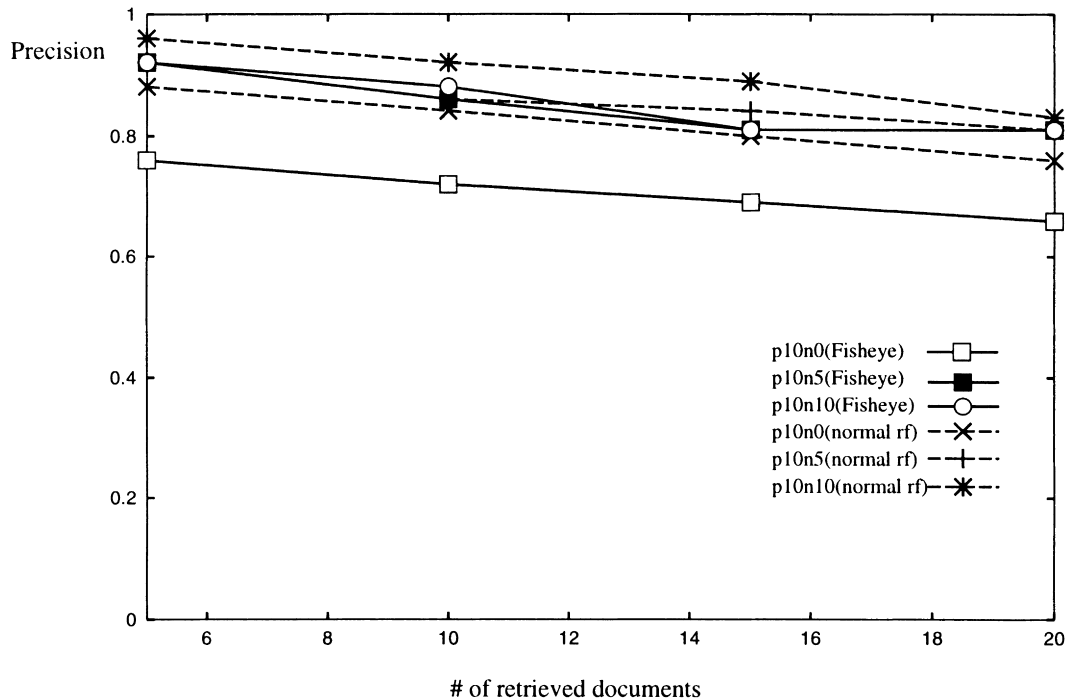


Fig. 3. Comparison of Fisheye Matching method (Fisheye) against the normal Relevance Feedback method (normal rf) when both positive and negative samples are inputted.

semantic group decreases as the number of positive samples increased. From this fact, we can conclude that the precision value is improved because the Fisheye Matching method achieves the following:

- The weights of words calculated by Eq. (8) can correspond to the users' viewpoint more precisely according to the number of positive samples.
- Semantic groups of appropriate size, which correspond to the concepts of not too general meanings, can be extracted.

After this experiment, we tried to improve the retrieving ability of the Fisheye Matching method further by removing the following two kinds of noise.

- The noise by using inappropriate semantic groups as features.
- The noise by existing inappropriate words in semantic groups.

To remove inappropriate semantic groups, we gave the system both positive and negative samples. Fig. 3 shows the precision value when both positive and negative samples are given. The number of positive samples were constant (10 articles), and the number of negative samples were varied from 0 to 10 articles. As in the case of experiments with positive samples only, five trials are done for each experimental condition, and the lines in this figure indicate the average score of each condition. The same set of positive and negative samples is given for all methods with the same condition.

In Fig. 3, although the precision value of the normal rf method can be improved with negative samples as in the case of the Fisheye Matching method, the difference between the normal rf and the Fisheye Matching methods becomes small compared with the case where only positive samples are given (Fig. 2). Therefore, the precision value of the Fisheye Matching method cannot be improved only by the property of VSM, but also by a property specific to the Fisheye Matching method. This property is shown in Table 1: the more the number of negative samples increased, both the number of semantic groups found and the average number of words that belong to each semantic group decreased. (Notice that the number of found semantic groups increased in the case of only positive samples.) From this fact we can conclude that giving negative samples can improve the retrieving ability of the Fisheye Matching method by avoiding inappropriate semantic groups being found and used.

In this paper, a word is considered as inappropriate if it does not exist in the sentence actually. For example, the stemming process may generate inappropriate words, because a certain stemmed word can be generated from different words. Furthermore, some phrases often have different meaning from its constituent words. In that case, the words in the phrase are also regarded as inappropriate, because they do not exist in the sentence as it is, in other words, their meaning never exists in the sentence.

For removing inappropriate words, the use of a morphological analyzer to extract words from documents will be effective. As for this expectation, we did the same

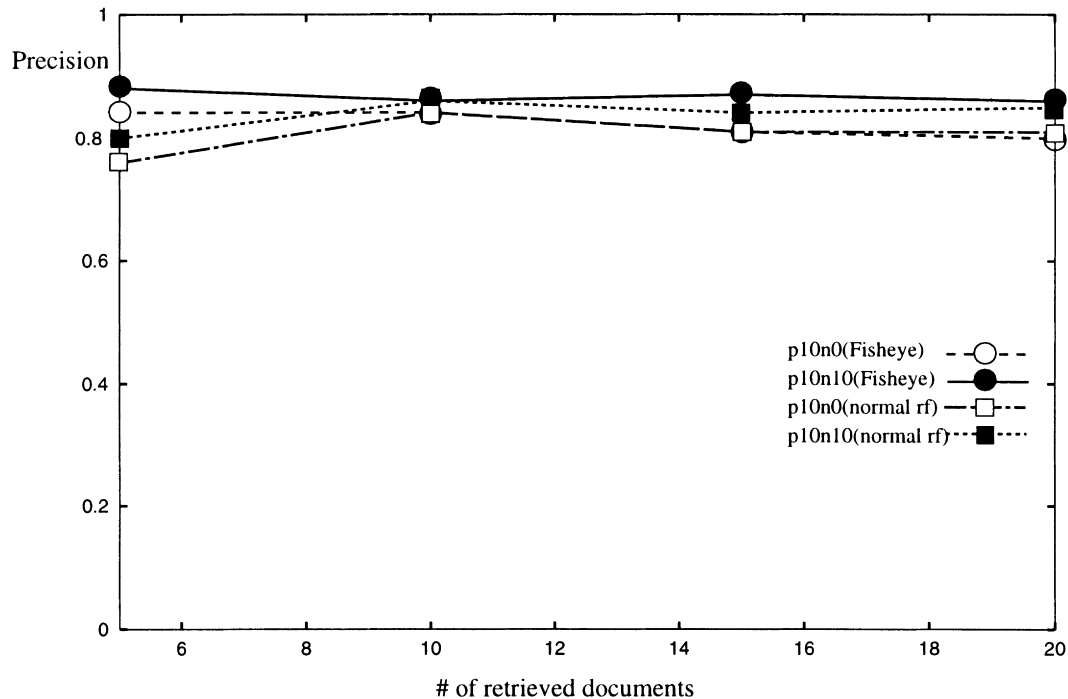


Fig. 4. Comparison of Fisheye Matching method (Fisheye) against the normal Relevance Feedback method (normal rf) with Japanese news articles.

experiment as reported above with the news articles written in Japanese using the morphological analyzer “chasen”,<sup>6</sup> and the results are shown in the Fig. 4. As in the case of experiments with English articles, five trials are done for each experimental condition, and the lines in this figure indicate the average score of each condition. The same set of positive and negative samples is given for all methods with the same condition.

It is observed from the Fig. 4 that the Fisheye Matching method can sometimes achieve a higher precision value than the Relevance Feedback method under the same conditions.

#### 4.2. Examine expressing users' viewpoint ability

Table 2 shows examples of the semantic groups that are frequently extracted as features reflecting the viewpoints during the experiments with English articles.<sup>7</sup> From this table, we can also confirm that appropriate semantic groups could be found by the algorithms described in Section 3.2. Furthermore, supplying both heading information and words that belong to them will be useful for the users to grasp their current viewpoints. We claim that this ability is one of the most valuable merits of the Fisheye Matching method, especially when applied to a document ordering support tool.

<sup>6</sup> <http://cactus.aist-nara.ac.jp/lab/nlt/chasen.html>.

<sup>7</sup> It is noticed that the EDR dictionary shares the concept structure between English and Japanese word dictionary, and it is confirmed that those concepts shown in Table 2 are also selected as viewpoints in the experiments of Fig. 4.

## 5. Related work

There have been various researches on visualizing the users' viewpoint, most of which are about creative thinking. It is a major way to express the users' viewpoint to place objects (document and/or keyword) on a display with a spring model method [5], or some statistical methods like dual scaling method [3,4]. Strictly speaking, it requires hyper-dimensional space to express relations among many objects precisely. To place objects on display (two-dimensional), the distance between documents on hyper-dimensional space has to be translated into that of two-dimensional space. Both the dual scaling and the spring model methods place objects on a display by (locally) minimizing the difference between the distance on hyper-dimensional space and that of two-dimensional space. Because the resulting placement may differ depending on the initial placement on a display, which was done by the user, it can express his viewpoint that was hidden in the initial placement.

Table 2  
Examples of extracted semantic groups related to medical topics

ID	Heading info	Group words(stemmed)
3f98b3	Value of health	Diseas, sickne, health,...
444506	Component of living body	Protei, immuno, choles,...
30f6da	Internal organs	Eye, heart, lung, knee,...
3f969e	Disease	Syndro, aids, cancer, cold,...
44479c	Medical supplies	Drug, medici, laxati, acid,...
30f6f7	Medical instruments	Bandag, cathet, glasse,...

However, the placement by these methods tends to be complicated, and it is hard for the users to understand their viewpoints as the number of objects increases. By contrast, the Fisheye Matching method can reflect the users' viewpoints not only by calculating similarity among documents based on the users' viewpoints, but also by supplying their viewpoints in a readable form; that is, both the heading information and the included words of the semantic groups (which reflect the users' viewpoints) are presented. In particular, it is remarkable that the supplied information about their viewpoints is independent of the number of objects.

As for the research on generating (selecting) features, [9] use the  $k$ -mean clustering method to divide words to  $k$  groups. The Latent Semantic Indexing (LSI) [10] employs a statistical approach, namely singular-value decomposition. Although these approaches are useful for reducing the dimensions of the vector space or resolving the vocabulary problem such as ambiguity, they use no background knowledge about grouping words, and the resulting groups often have the following problems: (1) the users cannot see what concepts these groups represent; and (2) these groups tend to be too large size, that is, they correspond to the concepts of too general meanings.

In case of the Fisheye Matching method, semantic groups, as the basic units of grouping words, are calculated from the EDR concept structure dictionary, which consists of human's common knowledge. Accordingly, the resulting semantic groups are easy to understand in their meaning, and serve the users to grasp their viewpoints.

## 6. Conclusion

We have described the Fisheye Matching method, which extends the normal VSM to make use of the users' viewpoints. The Fisheye Matching method copes with the users' viewpoints both by constructing the viewpoint-sensitive

vector space and by supplying the users with information about their viewpoints in a readable form.

We are currently developing a tool that assists the users to order documents with the Fisheye Matching method. So far, some people have used this tool and have given favorable comments. We will continue examining this tool, and report on this in the near future.

## References

- [1] Y. Takama, M. Ishizuka, Fish-eye matching: a document organizing function based on the extraction of user's viewpoint using concept structure, *Journal of the Japanese Society of Artificial Intelligence* 14 (1) (1999) 93–101 (in Japanese).
- [2] K. Sugiyama, K. Misue, I. Watanabe, K. Nitta, Y. Takeda, Emergent media environment for idea creation support, *Knowledge-Based Systems* 10 (1) (1997) 51–58.
- [3] M. Sugimoto, K. Hori, S. Ohsuga, A system to visualize different viewpoints for supporting researcher's creativity, *Knowledge-Based Systems* 9 (6) (1996) 369–376.
- [4] Y. Sumi, K. Nishimoto, K. Mase, Facilitating human communications in personalized information spaces, in: *AAAI-96 Workshop on Internet-based Information Systems*, 1996, pp. 123–129.
- [5] K. Takasugi, S. Kunifuji, An idea inspiration system using spring model, *Japanese Society of Artificial Intelligence* (1996) 34–39 (in Japanese, SIG Notes J-9602).
- [6] R. Armstrong, D. Freitag, T. Joachims, T. Mitchell, Webwatcher: a learning apprentice for the World Wide Web, in: *AAAI Spring Symposium on Information Gathering from Distributed, Heterogeneous Environments*, 1995.
- [7] M. Balabanovic, Y. Shoham, Learning information retrieval agents: experiments with automated Web browsing, in: *AAAI Spring Symposium on Information Gathering from Distributed, Heterogeneous Environments*, 1995.
- [8] W.W. Cohen, Y. Singer, Learning to query the web, in: *AAAI-96 Workshop on Internet-based Information Systems*, 1996, pp. 16–25.
- [9] M.R. Wulfekuhler, W.F. Punch, Finding salient features for personal web page categories, in: *The Sixth International WWW Conference*, <http://www.scope.gmd.de/info/www6/index.html>, 1997.
- [10] S.T. Dumais, G.W. Furnas, T.K. Landauer, Using latent semantic analysis to improve access to textual information. in: *CHI'88, Conference on Human Factors in Computing*, 1998, pp. 281–285.