Summarizing Search Results using PLSI

Jun Harashima* **and Sadao Kurohashi**
Graduate School of Informatics
Kyoto University
Yoshida-honmachi, Sakyo-ku,
Kyoto, 606-8501, Japan
{harashima,kuro}@nlp.kuee.kyoto-u.ac.jp

Abstract

In this paper, we investigate generating a set of query-focused summaries from search results. Since there may be many topics related to a given query in the search results, in order to summarize these results, they should first be classified into topics, and then each topic should be summarized individually. In this summarization process, two types of redundancies need to be reduced. First, each topic summary should not contain any redundancy (we refer to this problem as redundancy within a summary). Second, a topic summary should not be similar to any other topic summary (we refer to this problem as redundancy between summaries). In this paper, we focus on the document clustering process and the reduction of redundancy between summaries in the summarization process. We also propose a method using PLSI to summarize search results. Evaluation results confirm that our method performs well in classifying search results and reducing the redundancy between summaries.

1 Introduction

Currently, the World Wide Web contains vast amounts of information. To make efficient use of this information, search engines are indispensable. However, search engines generally return only a long list containing the title and a snippet of each of the retrieved documents. While such lists are effective for navigational queries, they are not helpful to users with informational queries. Some systems (e.g., Clusty[1]) present keywords related to a given query together with the search results. It is, however, difficult for users to understand the relation between the keywords and the query, as the keywords are merely words or phrases out of context. To solve this problem, we address the task of generating a set of query-focused summaries from search results to present information about a given query using natural sentences.

Since there are generally many topics related to a query in the search results, the task of summarizing these results is one of, so to speak, multi-topic multi-document summarization. Studies on multi-document summarization typically address summarizing documents related to a single topic (e.g., TAC[2]). However we need to address summarizing documents related to multiple topics when considering the summarization of search results.

To summarize documents containing multiple topics, we first need to classify them into topics. For example, if a set of documents related to *swine flu* contains topics such as the outbreaks of *swine flu*, the measures to treat *swine flu*, and so on, the documents should be divided into these topics and summarized individually. Note that a method for soft clustering should be employed in this process, as one document may belong to several topics.

*Research Fellow of the Japan Society for the Promotion of Science (JSPS)

[1] http://clusty.com/

[2] http://www.nist.gov/tac/

Proceedings of the Second Workshop on NLP Challenges in the Information Explosion Era (NLPIX 2010), pages 12–20,
Beijing, August 2010

In the summarization process, two types of redundancies need to be addressed. First, each topic summary should not contain any redundancy. We refer to this problem as redundancy within a summary. This problem is well known in the field of multi-document summarization (Mani, 2001) and several methods have been proposed to solve it, such as Maximum Marginal Relevance (MMR) (Goldstein et al., 2000) (Mori et al., 2004), using Integer Linear Programming (ILP) (Filatova and Hatzivassiloglou, 2004) (McDonald, 2007) (Takamura and Okumura, 2009), and so on.

Second, no topic summary should be similar to any of the other topic summaries. We refer to this problem as redundancy between summaries. For example, to summarize the above-mentioned documents related to *swine flu*, the summary for outbreaks should contain specific information about outbreaks, whereas the summary for measures should contain specific information about measures. This problem is characteristic of multi-topic multi-document summarization. Some methods have been proposed to generate topic summaries from documents (Radev and Fan, 2000) (Haghighi and Vanderwende, 2009), but to the best of our knowledge, the redundancy between summaries has not yet been addressed in any study.

In this paper, we focus on the document clustering process and the reduction of redundancy between summaries in the summarization process. Furthermore, we propose a method using PLSI (Hofmann, 1999) to summarize search results. In the proposed method, we employ PLSI to estimate the membership degree of each document to each topic, and then classify the search results into topics using this information. In the same way, we employ PLSI to estimate the membership degree of each keyword to each topic, and then extract the important sentences specific to each topic using this information in order to reduce the redundancy between summaries. The evaluation results show that our method performs well in classifying search results and successfully reduces the redundancy between summaries.

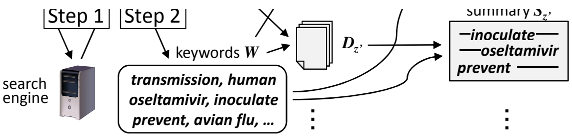


Figure 1: Overview of the proposed method.

2 Proposed Method

2.1 Overview

Figure 1 gives an overview of the proposed method, which comprises the following four steps.

Step 1. Acquisition of Search Results Using a search engine, obtain the search results for a given query.

Step 2. Keyword Extraction Extract the keywords related to the query from the search results using the method proposed by Shibata et al. (2009).

Step 3. Document Clustering Estimate the membership degree of each document to each topic using PLSI, and classify the search results into topics.

Step 4. Summarization For each topic, generate a summary by extracting the important sentences specific to each topic from each document cluster.

In the following subsections, we describe each step in detail.

2.2 Step 1. Acquisition of Search Results

First, we obtain the search results for a given query using a search engine. To be more precise, we obtain the top N' documents of the search engine results. Next, we remove those documents that should not be included in the summarization, such as link collections, using a simple filtering method. For example, we regard any document that has too many links as a link collection, and remove it.

In this paper, we write D to denote the search results after the filtering, and let $N = |D|$.

2.3 Step 2. Keyword Extraction

We extract the keywords related to a query from D using the method proposed by Shibata et al. (2009), which comprises the following four steps.

Step 2-1. Relevant Sentence Extraction For each document in D, extract the sentences containing the query and the sentences around the query as relevant sentences.

Step 2-2. Keyword Candidate Extraction For each relevant sentence, extract compound nouns and parenthetic strings as keyword candidates.

Step 2-3. Synonymous Candidate Unification Find the paraphrase pairs and the orthographic variant pairs in the keyword candidates, and merge them.

Step 2-4. Keyword Selection Score each keyword candidate, rank them, and select the best M as the keywords related to the query.

In this paper, we write W to denote the extracted keywords.

2.4 Step 3. Document Clustering

We classify D into topics using PLSI. In PLSI, a document d and a word w are assumed to be conditionally independent given a topic z, and the joint probability $p(d, w)$ is calculated as follows.

$$p(d,w) = \sum_z p(z)\ p(d|z)\ p(w|z) \tag{1}$$

$p(z)$, $p(d|z)$, and $p(w|z)$ are estimated by maximizing the log-likelihood function L, which is calculated as

$$L = \sum_d \sum_w freq(d,w) \log p(d,w), \tag{2}$$

where $freq(d, w)$ represents the frequency of word w in document d. L is maximized using the EM algorithm, in which the E-step and M-step are given below.

E-step

$$p(z|d,w) = \frac{p(z)\ p(d|z)\ p(w|z)}{\sum_{z'} p(z')\ p(d|z')\ p(w|z')} \tag{3}$$

M-step

$$p(z) = \frac{\sum_d \sum_w freq(d,w)\ p(z|d,w)}{\sum_d \sum_w freq(d,w)} \tag{4}$$

$$p(d|z) = \frac{\sum_w freq(d,w)\ p(z|d,w)}{\sum_{d'} \sum_w freq(d',w)\ p(z|d',w)} \tag{5}$$

$$p(w|z) = \frac{\sum_d freq(d,w)\ p(z|d,w)}{\sum_d \sum_{w'} freq(d,w')\ p(z|d,w')} \tag{6}$$

The EM algorithm iterates through these steps until convergence.

First, we give PLSI the number of topics K, the search results D, and the keywords W as input, and estimate $p(z)$, $p(d|z)$, and $p(w|z)$, where z is a topic related to the query, d is a document in D, and w is a keyword in W. There is, however, no way of knowing the value of K; that is, we do not know in advance how many topics related to the query there are in the search results. Hence, we perform PLSI for several values of K, and select the K that has the minimum Akaike Information Criterion (AIC) (Akaike, 1974), calculated as follows.

$$AIC = -2L + 2K(N+M) \tag{7}$$

Furthermore, we select $p(z)$, $p(d|z)$, and $p(w|z)$ estimated using the selected K as the result of PLSI.

Next, we calculate the membership degree of each document to each topic. The membership degree of document d to topic z, denoted $p(z|d)$, is calculated as

$$p(z|d) = \frac{p(d|z)\ p(z)}{\sum_{z'} p(d|z')}. \tag{8}$$

Finally, for each topic, we collect those documents whose membership degree to the topic is larger than the threshold α. If there is a document whose membership degree to multiple topics is larger than the threshold, we classify the document into each topic.

In this paper, D_z denotes the documents classified into topic z.

2.5 Step 4. Summarization

For each topic, we extract the important sentences specific to that topic from each document

Figure 2: Algorithm for summarization.

```
Input: A set of $K$ document clusters $\{D_z\}(z \in Z)$
Output: A set of $K$ summaries $\{S_z\}(z \in Z)$
Procedure:
1: for all $z \in Z$
2:     while $|S_z| < num(z)$
3:         for all $s \in D_z$
4:             calculate $s\_score(z, s, S_z)$
5:         $s_{max} = argmax_{s \in D_z \setminus S_z}\ s\_score(z, s, S_z)$
6:         $S_z = S_z \cup \{s_{max}\}$
7:     return $S_z$
```

cluster. Figure 2 gives the algorithm for summarization. When we generate the summary S_z for topic z, we calculate the importance of sentence s to topic z, denoted as $s\_score(z, s, S_z)$, for each sentence in D_z (lines 3-4). Then we extract the sentence s_{max} with the maximum importance as an important sentence, and include s_{max} in S_z (lines 5-6). When we extract the next important sentence, we recalculate the importance $s\_score(z, s, S_z)$ for each sentence in D_z except the sentence in S_z (lines 3-4). Then we extract the sentence s_{max} with the maximum importance as an important sentence, and add s_{max} to S_z (lines 5-6). We continue this process until the number of important sentences composing the summary, denoted $|S_z|$, reaches the number of important sentences extracted for topic z, denoted $num(z)$ (line 2).

$s\_score(z, s, S_z)$ is calculated as follows:

$$s\_score(z, s, S_z) = \sum_{w \in W_s} \Big(w\_score(z, w) \times c\_score(w, S_z, s) \Big) \tag{9}$$

where W_s represents the keywords in sentence s.

$w\_score(z, w)$ is a function to reduce the redundancy between summaries, and represents the importance of keyword w to topic z. We can use the probability of w given z, denoted $p(w|z)$, as the $w\_score(z, w)$. This approach fails, however, because if there are keywords with a high probability in both topic z and another topic z', the sentences containing such keywords are extracted as the important sentences in both topics, and it follows that the generated summaries will contain redundancy. To solve this problem, we use the membership degree of keyword w to topic z, denoted $p(z|w)$, as $w\_score(z, w)$. We use $p(z)$ and $p(w|z)$ estimated using PLSI in Section 2.4 to calculate $p(z|w)$.

$$p(z|w) = \frac{p(w|z)\ p(z)}{\sum_{z'} p(w|z')} \tag{10}$$

Keywords with high probability in several topics should have a low membership degree to each topic. Thus, using $p(z|w)$ as the $w\_score(z, w)$ prevents extracting sentences containing such keywords as important sentences, and it follows that the similarity between the summaries is reduced. Furthermore, the keywords which are specific to a topic are supposed to have a high membership degree to that topic. Thus, using $p(z|w)$ as $w\_score(z, w)$ makes it easier to extract sentences containing such keywords as important sentences, and with the result that each summary is specific to the particular topic.

$c\_score(w, S_z, s)$ is a function to reduce the redundancy within a summary, and represents the importance of a keyword w in a sentence s under the condition that there is a set of extracted important sentences S_z. The value of $c\_score(w, S_z, s)$ is determined mainly by whether or not w is contained in S_z. Table 1 gives the values of $c\_score(w, S_z, s)$. For example, if w is contained in S_z, we set $c\_score(w, S_z, s) = 0$, else we set $c\_score(w, S_z, s) = 1$. In this way, we can extract the sentences containing the keywords that are not contained in S_z as important sentences, and reduce the redundancy within the summary. Note that we make some exceptions to generate a coherent summary. For example, even if w is contained in S_z, we set $c\_score(w, S_z, s) = 2$ as long as w is the subject of s. In the same way, even if w is not contained in S_z, we set $c\_score(w, S_z, s) = -2$ as long as w is the subject of s. These values for $c\_score(w, S_z, s)$ are empirically determined.

Table 1: Values of $c\_score(w, S_z, s)$.

	w is contained in S_z	w is not contained in S_z
w is the subject of s	2	-2
otherwise	0	1

Finally, using $p(z)$ we determine the number of important sentences extracted for topic z, denoted as $num(z)$.

$$num(z) = \begin{cases} \lfloor I \times p(z) \rfloor & (p(z) \geq \beta) \\ I_{min} & (p(z) < \beta) \end{cases} \quad (11)$$

where I represents the parameter that controls the total number of important sentences extracted for each topic. The higher the probability a topic has, the more important sentences we extract. Note that no matter how low $p(z)$ is, we extract at least I_{min} important sentences.

3 Experiments

3.1 Overview

To evaluate the proposed method, we recruited 48 subjects, mainly IT workers, and asked them to fill in a questionnaire. We prepared a system implemented according to our method, and asked the subjects to use our system to evaluate the following four aspects of our method.

- Validity of the number of topics
- Precision of document clustering
- Degree of reduction in redundancy between summaries
- Effectiveness of the method for presenting information through summaries

We allowed the subjects to create arbitrary queries for our system.

3.2 System

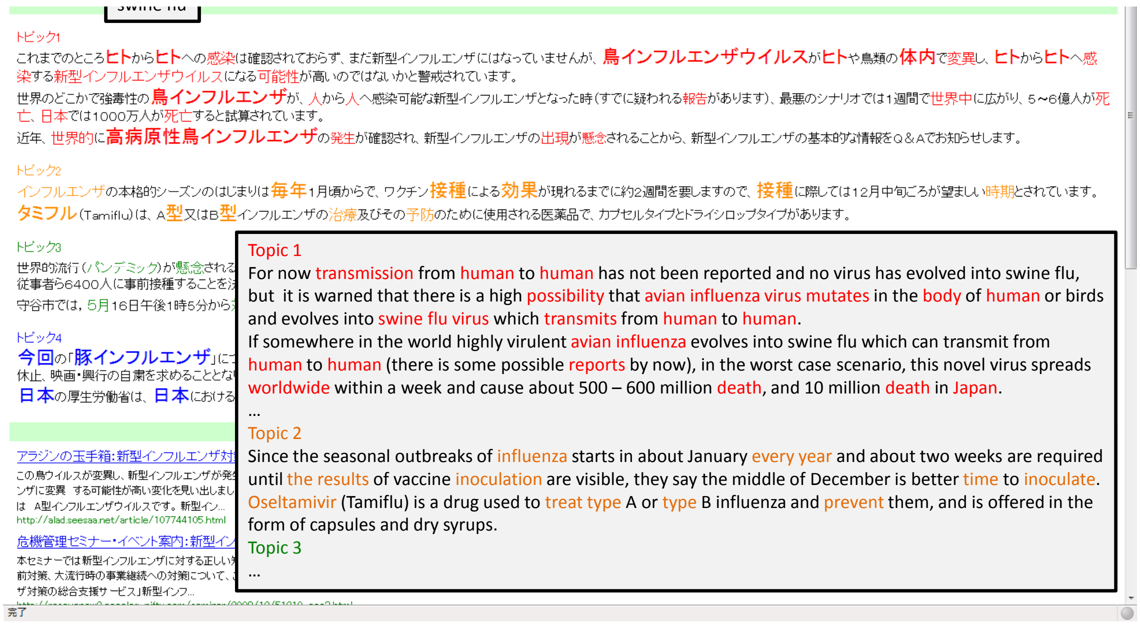
Figure 3 shows the system results for the query *swine flu*. Our system presents a separate summary for each topic related to a given query. In Fig.3, the colored words in the summaries are keywords specific to each topic. If a user clicks on a keyword, the system presents a list of documents containing that keyword at the bottom of the browser.

The configuration of our system was as follows. In the acquisition process, the system obtained the search results for a given query using the search engine TSUBAKI (Shinzato et al., 2008b). Setting $N' = 1,000$, we obtained the top $1,000$ documents in the search results for the query. In the keyword extraction process, we set $M = 100$, and extracted 100 keywords related to the query from the search results. In the document clustering process, we performed PLSI for $K = 3, 4, 5$, and selected the K with the minimum AIC. We set the initial value of $p(z) = 1/K$, and the initial values of $p(d|z)$ and $p(w|z)$ to random values. The EM algorithm continued until the increase in L reached just below 1 to achieve convergence. We set $\alpha = 1/K$. In the summarization process, we set $I = 10$, since the number of important sentences able to be presented in a browser is about 10. We set $I_{min} = 2$ and $\beta = 0.2$, and extracted at least two important sentences, even if $p(z)$ was very low.

3.3 Validity of the Number of Topics

First, we investigated how well the proposed method determined the number of topics. In our method, the number is determined using AIC. Ideally, we should have manually counted the topics in the search results, and compared this with the number determined using AIC. It was, however, difficult to count the topics, because the search results contained $1,000$ documents. Furthermore, it was impossible to count the number of topics for each query given by each subject. Thus, in this investigation, we simply asked the subjects whether they felt the number of topic summaries presented to them was appropriate or not, and investigated our method in terms of usability.

Table 2 gives the results. According to Table 2, 60.4% of the subjects agreed that the number of topic summaries presented by our system was acceptable. The average of the number of topics determined by our method was 3.18 per 1 query. On the other hand, 33.3% of the subjects said the number of topic summaries was too low or somewhat too low. According to these results, it seems that users are satisfied with the system presenting about 3 or 4 topic summaries, and our method determined the desirable number of topics in terms of usability.



トピック1

これまでのところヒトからヒトへの感染は確認されておらず、まだ新型インフルエンザにはなっていませんが、鳥インフルエンザウイルスがヒトや鳥類の体内で変異し、ヒトからヒトへ感染する新型インフルエンザウイルスになる可能性が高いのではないかと警戒されています。

世界のどこかで強毒性の鳥インフルエンザが、人から人へ感染可能な新型インフルエンザとなった時（すでに疑われる報告があります）、最悪のシナリオでは1週間で世界中に広がり、5～6億人が死亡、日本では1000万人が死亡すると試算されています。

近年、世界的に高病原性鳥インフルエンザの発生が確認され、新型インフルエンザの出現が懸念されることから、新型インフルエンザの基本的な情報をＱ＆Ａでお知らせします。

トピック2

インフルエンザの本格的シーズンのはじまりは毎年1月頃からで、ワクチン接種による効果が現れるまでに約2週間を要しますので、接種に際しては12月中旬ごろが望ましい時期とされています。

タミフル（Tamiflu）は、A型又はB型インフルエンザの治療及びその予防のために使用される医薬品で、カプセルタイプとドライシロップタイプがあります。

トピック3

世界的流行（パンデミック）が懸念される
従事者ら6400人に事前接種することを
守谷市では，5月16日午後1時5分から

トピック4

今回の「豚インフルエンザ」に
休止、映画・興行の自粛を求めることとな
日本の厚生労働省は、日本における

アラジンの玉手箱：新型インフルエンザ対
この鳥ウイルスが変異し、新型インフルエンザが発
ンザに変異　する可能性が高い変化を見い出しまし
は　A型インフルエンザウイルスです。新型イン…
http://alad.seesaa.net/article/107744105.html

危機管理セミナー・イベント案内：新型イン
本セミナーでは新型インフルエンザに対する正しい
前対策、大流行時の事業継続への対策について、
ザ対策の総合支援サービス」新型インフ…

完了

Topic 1
For now transmission from human to human has not been reported and no virus has evolved into swine flu, but it is warned that there is a high possibility that avian influenza virus mutates in the body of human or birds and evolves into swine flu virus which transmits from human to human.
If somewhere in the world highly virulent avian influenza evolves into swine flu which can transmit from human to human (there is some possible reports by now), in the worst case scenario, this novel virus spreads worldwide within a week and cause about 500 – 600 million death, and 10 million death in Japan.
…
Topic 2
Since the seasonal outbreaks of influenza starts in about January every year and about two weeks are required until the results of vaccine inoculation are visible, they say the middle of December is better time to inoculate.
Oseltamivir (Tamiflu) is a drug used to treat type A or type B influenza and prevent them, and is offered in the form of capsules and dry syrups.
Topic 3
…

Figure 3: System results for the query *swine flu*.

Table 2: Validity of the number of topics.

options	# subjects	(%)
(a) definitely too many	0	(0.0)
(b) somewhat too many	3	(6.3)
(c) acceptable	29	(60.4)
(d) somewhat too few	11	(22.9)
(e) definitely too few	5	(10.4)

3.4 Precision of Document Clustering

Second, we investigated how precisely the proposed method classified the search results into topics. To be more precise, we evaluated the reliability of the membership degree $p(z|d)$ used in the document clustering process. It is generally difficult to evaluate clustering methods. In our case, we did not have any correct data and could not even create these since, as mentioned previously, the number of topics is not known. Furthermore, it is not possible to classify by hand search results containing $1,000$ documents. Consequently, we did not evaluate our method directly by comparing correct data with the clustering result from our method, but instead evaluated it indirectly by investigating the reliability of the membership degree $p(z|d)$ used in the document clustering process.

The evaluation process was as follows. First, we presented the subjects with a document d, which was estimated by our system to have a high membership degree to a topic z. Strictly speaking, we selected as d, a document with a membership degree of about 0.9. Next, we presented two documents to the subjects. One was a document d' whose membership degree to z was also about 0.9, and another was a document d'' whose membership degree to z was about 0.1. Finally, we asked them which document was more similar to d[3].

Table 3 gives the results. According to this table, 60.5% of the subjects said d' was more similar or somewhat more similar. On the other hand, only 12.6% of the subjects said d'' was more similar or somewhat more similar. We see from these results that the ability to recognize topics in our system is in agreement to some extent with

[3] Naturally, we did not tell them that d' had a similar membership degree to d, whereas d'' did not.

Table 3: Precision of the estimation $p(z|d)$.

options	# subjects	(%)
(a) d' is definitely more similar	14	(29.2)
(b) d' is somewhat more similar	15	(31.3)
(c) undecided	13	(27.1)
(d) d'' is somewhat more similar	3	(6.3)
(e) d'' is definitely more similar	3	(6.3)

the subjects' ability for recognizing topics; that is, our method was able to estimate a reliable membership degree $p(z|d)$. Thus, it seems that our method using $p(z|d)$ is able to classify search results into topics to some extent.

3.5 Degree of Reduction in Redundancy between Summaries

Third, we investigated how well the proposed method reduced the redundancy between summaries. To be more precise, we used three measures as $w\_score(z, w)$ to generate summaries and investigated which measure generated the least redundant summaries. Generally, methods for reducing redundancy are evaluated using ROUGE (Lin, 2004), BE (Hovy et al., 2005), or Pyramid (Nenkova and Passonneau, 2004). However, the use of these methods require that ideal summaries are created by humans, and this was not possible for the same reason as mentioned previously. Thus, we did not perform a direct evaluation using the methods such as ROUGE, but instead evaluated how well our method performed in reducing redundancy between summaries using the membership degree $p(z|w)$ as $w\_score(z, w)$.

The evaluation process was as follows. We used three measures as $w\_score(z, w)$, and generated three sets of summaries.

Summaries A This set of summaries was generated using $dfidf(w)$ as $w\_score(z, w)$, with $dfidf(w)$ calculated as $ldf(w) \times log(100\, million/gdf(w))$, $ldf(w)$ representing the document frequency of keyword w in the search results, and $gdf(w)$ representing the document frequency of keyword w in the TSUBAKI document collection (Shinzato et al., 2008a) comprising about 100 million documents.

Table 4: Comparison of $dfidf(w)$, $p(w|z)$ and $p(z|w)$.

options	# subjects	(%)
(a) B is definitely less redundant	5	(10.4)
(b) B is somewhat less redundant	16	(33.3)
(c) undecided	15	(31.3)
(d) A is somewhat less redundant	6	(12.5)
(e) A is definitely less redundant	6	(12.5)
options	# subjects	(%)
(a) C is definitely less redundant	16	(33.3)
(b) C is somewhat less redundant	14	(29.2)
(c) undecided	6	(12.5)
(d) A is somewhat less redundant	8	(16.7)
(e) A is definitely less redundant	4	(8.3)
options	# subjects	(%)
(a) C is definitely less redundant	15	(31.3)
(b) C is somewhat less redundant	8	(16.7)
(c) undecided	10	(20.8)
(d) B is somewhat less redundant	6	(12.5)
(e) B is definitely less redundant	9	(18.8)

Summaries B This set of summaries was generated using $p(w|z)$ as $w\_score(z, w)$.

Summaries C This set of summaries was generated using $p(z|w)$ as $w\_score(z, w)$.

We then presented the subjects with three pairs of summaries, namely a pair from A and B, a pair from A and C, and a pair from B and C, and asked them which summaries in each pair was less redundant[4].

The results are given in Tables 4. Firstly, according to the comparison of A and B and the comparison of A and C, A was more redundant than B and C. The value of $dfidf(w)$ to keyword w was the same for all topics. Thus, using $dfidf(w)$ as $w\_score(z, w)$ made summaries redundant, as each summary tended to contain the same keywords with high $dfidf(w)$. On the other hand, as the value of $p(w|z)$ and $p(z|w)$ were dependent on the topic, the summaries generated using these measures were less redundant.

Second, the comparison of B and C shows that 48.0% of the subjects considered C to be less redundant or somewhat less redundant. $p(w|z)$ was a better measure than $dfidf(w)$, but even using $p(w|z)$ generated redundancy between sum-

[4] Naturally, we did not tell them which summaries were generated using which measures

Table 5: Comparison of summaries and keywords.

options	# subjects	(%)
(a) X is definitely more helpful	25	(52.1)
(b) X is somewhat more helpful	10	(20.8)
(c) undecided	3	(6.3)
(d) Y is somewhat more helpful	8	(16.7)
(e) Y is definitely more helpful	2	(4.2)

maries. Because common keywords to a query have high $p(w|z)$ for several topics, sentences containing these keywords were extracted as the important sentences for those topics, and thus the summaries were similar to one another. On the other hand, the keywords' value for $p(z|w)$ was low, allowing us to extract the important sentences specific to each topic using $p(z|w)$ as $w\_score(z, w)$, thereby reducing the redundancy between summaries.

3.6 Effectiveness of the Method for Presenting Information Using Summaries

We also investigated the effectiveness of the method for presenting information through summaries. We asked the subjects to compare two different ways of presenting information and to judge which way was more effective in terms of usefulness for collecting information about a query. One of the methods presented the search results with topic summaries generated by our system (method X), and while the another method presented the search results with the keywords included in each topic summary (method Y).

Table 5 gives the results. 72.9% of the subjects considered the method using summaries to be more effective or somewhat more effective. From these results, it appears that the method of presenting information through summaries is effective in terms of usefulness for collecting information about a query.

4 Conclusion

In this paper, we focused on the task of generating a set of query-focused summaries from search results. To summarize the search results for a given query, a process of classifying them into topics related to the query was needed. In the proposed method, we employed PLSI to estimate the membership degree of each document to each topic, and then classified search results into topics using this metric. The evaluation results showed that our method estimated reliable degrees of membership. Thus, it seems that our method is able to some extent to classify search results into topics.

In the summarization process, redundancy within a summary and redundancy between summaries needs to be reduced. In this paper, we focused on the reduction of the latter redundancy. Our method made use of PLSI to estimate the membership degree of each keyword to each topic, and then extracted the important sentences specific to each topic using this metric. The evaluation results showed that our method was able to reduce the redundancy between summaries using the membership degree.

In future, we will investigate the use of more sophisticated topic models. Although our method detected the topics related to a query using a simple topic model (i.e., PLSI), we believe that more sophisticated topic models such as LDA (Blei et al., 2003) allow us to improve our method.

References

Akaike, Hirotugu. 1974. A new look at the statistical model identification. *IEEE Transactions on Automation Control*, 19(6):716–723.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022.

Filatova, Elena and Vasileios Hatzivassiloglou. 2004. A formal model for information selection in multi-sentence text extraction. In *Proceedings of COLING 2004*, pages 397–403.

Goldstein, Jade, Vibhu Mittal, Jaime Carbonell, and Mark Kantrowitz. 2000. Multi-document summarization by sentence extraction. In *Proceedings of ANLP/NAACL 2000 Workshop on Automatic Summarization*, pages 40–48.

Haghighi, Aria and Lucy Vanderwende. 2009. Exploring content models for multi-document summarization. In *Proceedings of HLT-NAACL 2009*, pages 362–370.

Hofmann, Thomas. 1999. Probabilistic latent semantic indexing. In *Proceedings of SIGIR 1999*, pages 50–57.

Hovy, Eduard, Chin-Yew Lin, and Liang Zhou. 2005. Evaluating duc 2005 using basic elements. In *Proceedings of DUC 2005*.

Lin, Chin-Yew. 2004. Rouge: A package for automatic evaluation of summaries. In *Proceedings of ACL 2004 Workshop on Text Summarization Branches Out*, pages 74–81.

Mani, Inderjeet. 2001. *Automatic Summarization*. John Benjamins Publishing Company.

McDonald, Ryan. 2007. A study of global inference algorithms in multi-document summarization. In *Proceedings of ECIR 2007*, pages 557–564.

Mori, Tatsunori, Masanori Nozawa, and Yoshiaki Asada. 2004. Multi-answer-focused multi-document summarization using a question-answering engine. In *Proceedings of COLING 2004*, pages 439–445.

Nenkova, Ani and Rebecca Passonneau. 2004. Evaluating content selection in summarization: The pyramid method. In *Proceedings of NAACL-HLT 2004*.

Radev, Dragomir R. and Weiguo Fan. 2000. Automatic summarization of search engine hit lists. In *Proceedings of ACL 2000 Workshop on Recent advances in NLP and IR*, pages 1361–1374.

Shibata, Tomohide, Yasuo Bamba, Keiji Shinzato, and Sadao Kurohashi. 2009. Web information organization using keyword distillation based clustering. In *Proceedings of WI 2009*, pages 325–330.

Shinzato, Keiji, Daisuke Kawahara, Chikara Hashimoto, and Sadao Kurohashi. 2008a. A large-scale web data collection as a natural language processing infrastructure. In *Proceedings of LREC 2008*, pages 2236–2241.

Shinzato, Keiji, Tomohide Shibata, Daisuke Kawahara, Chikara Hashimoto, and Sadao Kurohashi. 2008b. TSUBAKI: An Open Search Engine Infrastructure for Developing New Information Access Methodology. In *Proceedings of IJCNLP 2008*, pages 189–196.

Takamura, Hiroya and Manabu Okumura. 2009. Text summarization model based on maximum coverage problem and its variant. In *Proceedings of EACL 2009*, pages 781–789.