

# A Study for Documents Summarization based on Personal Annotation

**Haiqin Zhang**  
University of Science and  
Technology of China  
face@mail.ustc.edu.  
cn

**Zheng Chen Wei-ying Ma**  
Microsoft Research Asia  
zhengc@microsoft.com  
wyma@microsoft.com

**Qingsheng Cai**  
University of Science and  
Technology of China  
qscai@ustc.edu.cn

## Abstract

For one document, current summarization systems produce a uniform version of summary for all users. Personalized summarizations are necessary in order to represent users' preferences and interests. Annotation is getting important for document sharing and collaborative filtering, which in fact record users' dynamic behaviors compared to traditional steady profiles. In this paper we introduce a new summarization system based on users' annotations. Annotations and their contexts are extracted to represent features of sentences, which are given different weights for representation of the document. Our system produces two versions of summaries for each document: generic summary without considering annotations and annotation-based summary. Since annotation is a kind of personal data, annotation-based summary is tailored to user's interests to some extent. We show by experiments that annotations can help a lot in improving summarization performance compared to no annotation consideration. At the same time, we make an extensive study on users' annotating behaviors and annotations distribution, and propose a variety of techniques to evaluate the relationships between annotations and summaries, such as how the number of annotations affects the summarizing performance. A study about collaborative filtering is also made to evaluate the summarization based on annotations of similar users.

## 1 Introduction

As information explodes in the Internet, it's hard for users to read through all the published materials potentially interesting. Therefore, it is of great help to present them in a condensed way, i.e. using extracts or abstracts that generalize the content of an article. Text summarization is the process of distilling the most important information from a source (or sources) to produce an abridged version for a particular user (or users) and task (or tasks) (Mani and Maybury, 1999). Summary can help the user quickly get a general idea about the article and decide if it deserves more detailed attention.

The problem with present summarization systems is that they produce one uniform summary for a given document without considering users' opinions on it, while different users may have different perspectives on the same text, thus need a different summary. A good summary should change corresponding to interests and preferences of its reader. We refer to the adaptation of the summarization process to a particular user as personalized summarization. We wish to extract the kind of personalized summary in this paper.

Annotating is a common behavior while reading, since many users would like to make some remarks or to highlight some important parts of the text. So we think that, to some extent, annotations represent users' interests by recording some viewpoints of users on the documents, such as which part is important or interesting, which is not. With the rapidly technologies improvements on tablet-pc, it is possible to record these behaviors automatically. From another point, annotations can be seen as a kind of user feedback, but different from traditional explicit relevance feedback (Salton and Buckley, 1990) in the sense of that annotation can be collected automatically without user's consciousness. Since annotation is a kind of personal data, it is expected to help improve the personalization of present

*\*This work was done when the first author visited Microsoft Research Asia.*

document summarizers; also, intention of users can be learned through this interaction process.

Since a user may make annotations freely, not all of them are useful for representing his real preferences. Therefore it is helpful to find how some of the annotations affect on summarization performance, based on that, we make a study on evaluations about the number of annotations effective.

In spite of summarization based on the user's annotation himself, we want to do more when his annotation on current document is not available. Suppose he is new to the domain, but other users annotated on the document, we can resort to annotations of similar users to help his summarization. Study in such situation is thought of as collaboratively filtering which filter the useful information from similar users to help the summarization of current user. To our knowledge, no work has been done about such annotation studies.

The remainder of the paper is organized as follows. Section 2 introduces some related work in text summarization. Section 3 presents the new annotation based text summarization approach, including annotation, context and keywords extraction, and summary generation. Section 4 gives the metrics for evaluation of annotated summarization. Section 5 discusses the detailed experiments and evaluation of summarization, annotations and collaborative filtering. Section 6 concludes the paper with a simple summary and future work.

## 2 Related work

To summarize, is to reduce complexity of documents, hence, in length, while retaining some of the important information in the original documents. Titles, keywords, tables-of-contents and abstracts might all be considered as forms of summary; here, we consider summary as a set of sentences containing some of the essential information from the original document.

A lot of approaches were proposed in text summarization, such as word frequency based method (Luhn, 1958), cue phrase method (Edmundson, 1969), Position-based methods (Edmundson, 1969; Hovy and Lin, 1997; Teufel and Moens, 1997).

At the same time, some machine learning methods were used to integrate different clues in documents. Given a corpus and its predefined summaries as training set, it is to identify the relationships between documents and their summaries, the sentences which satisfy the rules are the ones to be extracted (Kupiec et al., 1995). Other machine learning methods perform sentence clustering based on a set of extracted features of sentences, and, choose a representative sentence from each cluster, and combine them into a summary according to their original order in the text (Nomoto and Matsumoto, 2001).

Most of the above techniques have the limitations we have mentioned at the beginning, they failed to supply a personalized summary which reflects the interests and preferences of different users.

There were some work based on annotation (Golovchinsky et al., 1999; Price et al., 1998), but they mainly focus on supplying an authoring tool which gives instructions on how to do annotation; or annotation identification and extraction which is difficult since annotation may be done freely and randomly; or annotation based query which aims at query expansion based on annotations. But rarely people think of using annotation for summarization. In fact, we only found one work about summarization based on annotation (Nagao and Hasida, 1998), but annotations there are defined on a complex set of GDA (Global Document Annotation) tags, which is an XML-based tag set, and allows machines to automatically infer underlying structures of documents by parsing, and authors of WWW files can annotate their documents by those tags, but they are not studied further about how to affect the summarization.

Since annotations reflect user's opinions, different users may have different annotations; thus summarizations based on annotations are tailored to users' interests to some extent. Therefore we will integrate annotations into our summarization framework, which is expected to supply personalized summaries for given users, and different from traditional uniform summary. Here we make an assumption that what are annotated is interesting or important compared to other parts of document, which is reasonable since this is a common view about why users make annotations.

In this paper, we mainly focus on the kind of annotations that are parts of the text in order to avoid complex manuscript recognition. Since we mainly consider the effect of annotations on the performance of summarization, annotations can only make sense when they are thought of as keywords. However, past experiments show that keywords method has a lower performance when working with other methods (Edmundson, 1969; Teufel and Moens, 1997), so the main approach we used here is based on key words frequency.

## 3 Annotation based summarization

**Annotation** is defined as a body of words marked among the text. It may be any word, phrase or sentences which readers may feel interesting or important. When we say "annotation", we mean its position and content, that is, where it is located and what texts it contains. Since users may annotate part of the important information, or the annotations may be incomplete, therefore, in spite of annotations themselves, we also need to consider "context" as a supplement to what users are interested in. For a particular annotation, **context** is defined the surrounding text of the annotation.

Since annotations contain a set of keywords, its significance can be identified when compared to other words of the text. For a given document, we first extract user's annotations and their contexts, and construct a new keywords set together with original keywords in the text, where annotations and contexts are given higher scores than others; Then we weight sentences according to keywords they contained and do summarization by selecting high-weighted ones.

### 3.1 Annotations & Context Extraction

A set of sentences is extracted from a given document. And for each annotation, we identify which sentence it is located at, as the **context** of the annotation. Then keywords are extracted from annotations and their contexts. An annotation may span through several sentences, a sentence may include several annotations, and an annotation may contain several keywords. For each sentence, we simplify the keywords extraction problem as identifying the annotations it contains. **Annotated sentences** are defined as those who contain annotations. The keywords occurring in annotations are called **annotated keywords (F1)**. Keywords occurring in annotated sentences are called **context keywords (F2)**. Frequencies  $f$  of repetitive keywords in F1 or F2 are accumulated. It's obvious that F1 is a subset of F2. Thus we get two keywords vector set:

Annotated keywords (F1):

$$F1 = \langle w_{1_i}, f_{1_i} \rangle, 1 \leq i \leq n_1, n_1 = \text{size}(F1)$$

Context keywords (F2):

$$F2 = \langle w_{2_i}, f_{2_i} \rangle, 1 \leq i \leq n_2, n_2 = \text{size}(F2)$$

### 3.2 Keywords Extraction

From the document, content words are stemmed from Porter's algorithm (Porter, 1980). Content keywords are referred to words whose frequencies are beyond a certain threshold and not occurring in stopping wordlist. Word frequencies are calculated by tf\*idf method (Salton and Buckley, 1988). After applying word occurrences statistics to full text; we get the vector set  $\langle w_i, f_i \rangle$ .

**Text keywords** are those occurred either in F2 or their frequency satisfies a given threshold  $\alpha$  ( $f_i > \alpha$ ). Annotated words are considered superiorly whatever frequency they occur originally, since users may be interested in some rare or "unknown" keywords in the documents, this kind of words should not be excluded beyond text keywords.

Text keywords (F0):

$$F0 = \langle w_{0_i}, f_{0_i} \rangle, 1 \leq i \leq n_0, n_0 = \text{size}(F0)$$

It is obvious that both F1 and F2 are subsets of F0.

Next in order to apply the emphasis of annotations and contexts on summarization, combination is per-

formed to integrate F0, F1 and F2. Since keywords in different sets may have different influences on summarization, some parameters are used to balance their effects respectively.

**Final keywords (F):**  $F = F0 + \beta F1 + \gamma F2$ ;

$$F = \langle w_i, f_i \rangle, 1 \leq i \leq n, n = \text{size}(F)$$

$\beta$  is annotation weight ( $\beta \geq 0$ ),  $\gamma$  is context weight ( $\gamma \geq 0$ ).  $\beta=0$  means considering no annotations.  $\gamma=0$  means considering no context.

### 3.3 Sentence Extraction

Sentences are weighted according to the keywords it contains:

$$W_s = \sum_{w_i \in S} f_i \text{ or } \frac{1}{|S|} \sum_{w_i \in S} f_i^2$$

$|S|$  is the length of a sentence, which means the keyword count it contains. Sentences are ranked by their weights, and then top scored sentences are selected as important ones and used to compose into a summary according to their original position.

## 4 Evaluation

The problem of evaluating text summarization is a quite deep one, and some problems remain concerning the appropriate methods and types of evaluation. There are a variety of possible bases for comparison of summarization performance, e.g., summary to source, system to manual summary. In general, methods for text summarization can be classified into two categories (Firmin and B, 1998; Mani and Maybury, 1999). The first is intrinsic evaluation, which judge the quality of the summarization directly based on analysis of the summary, including user judgments of fluency of the summary, coverage of the "key/essential ideas", or similarly to an "ideal" summary which is hard to establish. The other is extrinsic evaluation, which judge the quality of the summarization based on how it affects on the completion of other tasks, such as question answering and comprehension tasks.

Here we use intrinsic evaluation for our summarization performance. It is to compare the system summary with an ideal manual summary. Since we need to collect annotations for experimented documents, which require reading through the text, manual summaries can be made consequently after the reading.

The documents dataset to be evaluated are supplied with human annotations and summaries, which will be described in detail in the next section.

For one annotated document, our annotation based summarization (ABS) system produce two versions of summaries: **generic summary** without considering annotations, and **annotated summary** considering annotations. For evaluation, we made comparison between human-made summary and generic summary, and com-

comparisons between human-made summary and annotated summary. There are a lot of measures to make the comparisons (Firmin and B, 1998; Mani and Maybury, 1999), such as precision, recall, some of which will be used for our evaluation. Another measure we are interested in is the cosine similarity for two summaries, which is defined on keywords, and reflects the general similarity of two summaries in global distribution. For human-made summary  $S_0$  and summary  $S_i$  generated by ABS, summary similarity is as follows:

$$SIM(S_0, S_i) = \frac{\sum_{w_j \in S_0 \cap S_i} f_{0,j} \cdot f_{i,j}}{\sqrt{(\sum_{w_j \in S_0} f_{0,j}^2) \cdot (\sum_{w_j \in S_i} f_{i,j}^2)}}$$

$$Precision = \frac{Number\ of\ correct\ items\ retrieved}{Number\ of\ retrieved\ items}$$

$$Recall = \frac{Number\ of\ correct\ items\ retrieved}{Number\ of\ correct\ items}$$

Precision and recall are generally applied to sentences; in fact they can be applied to keywords too, which reflects the percentage of keywords correctly identified. Therefore, in spite of summary similarity, our measures for evaluation also include sentences precision, sentences recall, keywords precision and keywords recall. For keywords evaluation, a keyword is correct only if it occurs in human-made summary. For sentences evaluation, a sentence in summary is correct if it has as many possible keywords as in the corresponding sentence in the human-made summary, that is, their similarity (calculated same as summary similarity) is beyond a certain threshold. We use two types of sentences match in the experiments: one is perfect match, which means a sentence in summary is correct only if it occurs in manual summary; the other is conditional match, which means most concepts of the two sentences are correct, in this case the match similarity threshold is less than 1.

For a set of annotated documents, average values for the above five measures are calculated to show the general performance of the comparison.

## 5 Experiments

Since our approaches are based on annotated documents, we need to collect users' annotations for a set of documents. In the meantime, users are required to supply a summary consisting of several sentences that reflects the main ideas of the document. We supply an annotating tool called "Annotation" that can be used to highlight words, phrases, or sentences which users are interested in, and to select important sentences into summary.

The original data we used is from Yahoo news, which is interesting to most users. The data is composed of eight categories: business, sports, entertainment,

world, science, technology, politics, and oddly, with totally about 6000 documents. Documents are preprocessed and removed graphics and tags before experiments. We hired five students 10 days to annotate the documents, each student were supplied 20 percent of the documents. They are allowed to choose articles interesting to do the experiments. Users are told to make annotations freely and summaries which reflect the main ideas of the text. The process of annotating and summarizing for one user are independent, that is, they are done at different time. At last, we collect totally 1098 different annotated documents, each of which consists of a set of annotations and a human-made summary. The statistics for five persons is presented in table 1<sup>1</sup>, which shows that the average summary length (sentences number) is 6.11, and the average annotations number is 11.86.

	P1	P2	P3	P4	P5
DN	192	198	312	199	197
ASN	25.93	23.38	30.82	33.89	24.02
ASL	6.75	3.05	6.15	7.67	6.97
ANN	11.54	8.08	15.18	9.98	12.77

Table 1. A user study.

### 5.1 Summarization Evaluation

Since different users may have different annotation style, we separate the experiments for each individual's data. In experiments, the keyword threshold  $\alpha$  is set 2, which is reasonable that most keywords' frequency is at least 2. The threshold for summary similarity related to sentences precision, is 0.3 (which in fact means its square root 55% sentences are correct). The summarizer produce the same number of sentences as are in the corresponding manual summary, as in (Kupiec et al., 1995), therefore, precision and recall are the same for summaries sentences comparison.

First, we make experiments with different annotations and context weights. The results are presented in Table.2. The first column contains different combinations of annotation weight  $\beta$  and context weight  $\gamma$ , as described in Section 3.2. See <sup>2</sup> for the meanings of other columns. It is obvious that context can help to improve the summarization performance than no context consideration, so in our later experiments, we set the context weight  $\gamma=1$ , and annotation weight  $\beta=1$ .

<sup>1</sup> For ease of representation in table 1, "Pi" is for different users; "DN" is document number; "ASN" is average sentences number; "ASL" is average summary length; "ANN" is average annotation number.

<sup>2</sup> For two compared summaries, "SS" means summary similarity; "SP" means sentences precision for conditional match; "PP" means sentences precision for perfect match; "KP" means keywords precision; and "KR" means keywords recall.

$\beta, \gamma$	SS(%)	SP(%)	PP(%)	KP(%)	KR(%)
1, 0	73.79	60.35	52.88	59.25	75.23
1, 1	74.26	60.86	53.62	59.86	75.13
2, 0	74.04	60.67	53.21	59.70	74.90
2, 1	74.63	61.48	53.95	60.46	75.03

Table 2. Comparison of annotated summary with different parameters.

Next, we make formal experiments for generic summarization and annotation based summarization. Table 3 presents the average performance results for five users' data<sup>3</sup>. From the above table and figure, we found that annotation-based summarization is much better than generic summarization. The improvements are quite inspiring. In the case of user P4, the cosine similarity is increased by 10.1%; the sentences precision for conditional match is increased by 13.57%; precision for perfect match is increased by 17.59%; keywords precision is increased by 11.6%; keywords recall is increased by 13.18%, which shows that annotations can help a lot to improve the summarization performance.

		SS(%)	SP(%)	PP(%)	KP(%)	KR(%)
P	G	71.65	57.15	49.55	54.71	71.36
1	A	77.24	64.98	59.13	61.17	81.00
P	G	53.05	43.26	37.80	43.83	56.65
2	A	63.10	52.90	50.98	54.01	67.50
P	G	69.29	53.84	42.40	52.77	66.84
3	A	73.64	59.21	49.73	57.26	73.31
P	G	67.73	49.21	36.81	52.17	63.94
4	A	77.74	62.88	54.40	63.77	77.12
P	G	76.64	60.54	49.09	60.21	73.85
5	A	80.06	65.36	56.18	64.57	77.91

Table 3. Comparison of generic summary and annotated summary for 5 users' data.

Figure 1 shows the average performance comparison for total 1098 documents. Compared with generic summarization, cosine similarity of annotation based summarization is increased by 6.47%; sentences precision for conditional match is increased by 7.99%; precision for perfect match increased by 10.62%; keywords precision increased by 7.14%; keywords recall increased by 8.61%. The most significant is that the improvement of sentences precision for perfect match is higher than 10%, since perfect matches require two sentences totally matched, it is very important to gain this point, showing that in general annotations are able to contribute much to the performance of summarization. In Figure 1, "BA" means 'first Best' Annotated summary, which will be explained in the next subsection.

<sup>3</sup> In Table 3 and later figures, "G" means comparison of manual summary and generic summary, and "A" means comparison of manual summary and annotated summary.

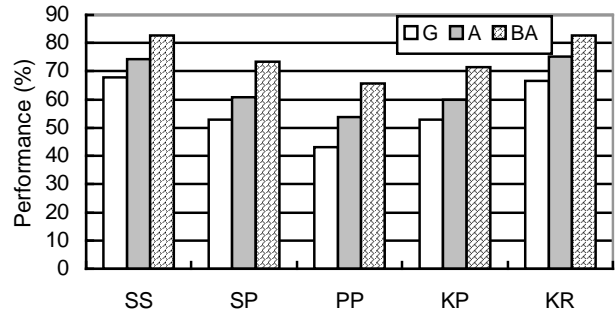


Figure 1. Performance comparison of generic summary, annotated summary and 'first best' summary.

## 5.2 Annotation Evaluation

In our last experiments, we found that the average annotation number was 11.86, which was much higher than summary length. We wonder whether there are some relations between the number of annotations and summarization performance. Thus we make such annotations evaluations to study how the number of annotations affects on summarization performance.

The first experiment is to find the best summary by selecting the first  $k$  ( $k \leq n$ ,  $n$  is total number of annotations) annotations, that is "first Best Annotated summary". In fact, from figure 1 we can see that when using all of the annotations, the performance falls about in the middle of generic summary and first Best Annotated summary (labeled by "BA"). However we found that, in some annotated documents, some of annotations are beyond the scope of the manual summary. This means that some of them are noisy to summarization; using all of them cannot reach the best performance; there must have a changing process of performance as more annotations are considered, which confirms us to explore the relationship between the annotation number and the summarization performance.

So the next experiment is to observe how the average summarization performance evolves as we select any of the  $k$  annotations. That is, for any annotation number  $k \leq n$ , to average the summarization performance by enumerating all possible annotations combinations. Figure 2 presents such a plot for a document "sports\_78.html", which indicates how the number of annotations affects summarization performance. It has totally 15 annotations, "0" stands for generic summary performance. When considering only one annotation, the performance drops a bit down (We found in this document some of the single annotations are far away from corresponding summary), but as more annotations considered, the performance begins to increase slowly and reaches the best at annotation number 12, then again begins to drop. For other documents, we find similar

situations that at beginning the performance increases along the number of annotations, but after it reaches a certain degree, the performance will fluctuate and sometimes drop slightly down. These are all true for our 5 evaluation measures, which are consistent and reasonable since too many annotations will introduce some degree of noise, which would bias the summarization process. We conclude from this evolving process that not all of the annotations are valuable for summarization.

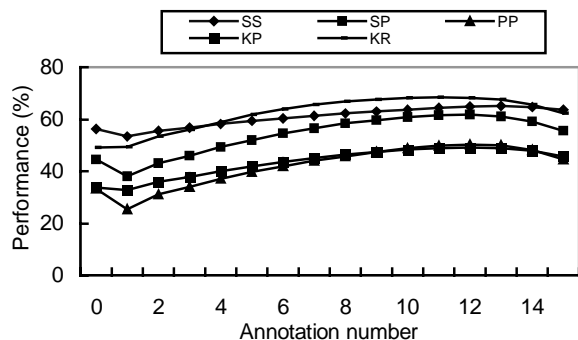


Figure 2. Performance along with the annotations number.

In figure 2, the best and worst performance point can be identified, for example, we get the best point at annotation number 12 and the worst point at annotation number 1. For a subset of 10 documents, summary similarity comparisons with generic, best, worst, and all annotations-based summarization are shown in figure 3.

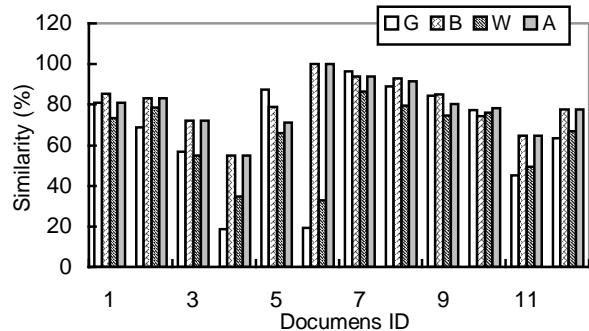


Figure 3. Comparisons of summaries with generic, best, worst, and all annotations.

Different documents have different annotations, which have different influences on summarization. In most cases, best summaries are better than all annotations-based summaries, which are better than generic and worst summaries. There is an exception in Figure 3 at Document 5, we found in this document some of the annotations are irrelevant to the summary. For example, in this document, percentage of summary annotated sentences in annotated sentences is 28.57%; and per-

centage of summary annotated keywords in summary keywords is only 17.19%. While for Document 6, the corresponding values are 50.00% and 32.73%. This indicates that the user's interests drifted as he read the document.

When annotations are consistent with users' summaries, they help to improve the personalization of summarization, the more the better, otherwise, when the annotations are far from users' summaries, the influence of annotations may be negative, for example some annotations subset make the performance worse. But generally the former has a much larger proportion than the latter. Along the documents in figure 3, the main trend is that annotation-based summaries are better than summaries with no annotations consideration; the average improvement for 12 documents is 13.49%.

### 5.3 Collaborative Filtering

Another part of our experiments is about collaborative filtering, which identifies how much effect one user's annotation is on others' summaries. To do that, we additionally invited the previous 5 users to annotate and summarize 100 common documents. After removing some bad data, we got totally 91 common documents at last. For each user's data, we will find whether it is helpful when his annotations are applied on other users' summarizations. Figure 4 presents the contributions of user P3's annotation on the total five persons' summarization. Intuitively, P3's annotations are most helpful to P3's summaries; they also have some contributions to P2, P4 and P1's according to importance, but make negative effects on P5's summarization. This indicates that P3's interest is most close to P2's, but far from P5's. This is possibly understandable since different users have different preferences for a document, thus their annotations style may have significant variances.

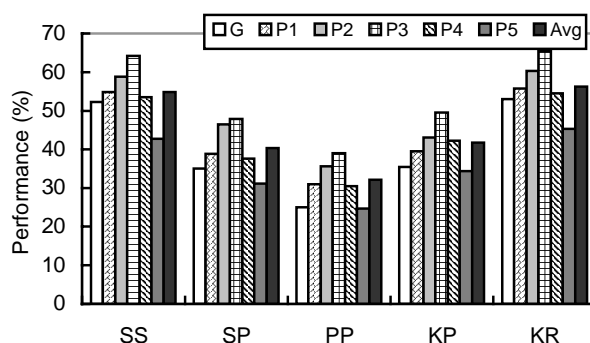


Figure 4. Contributions of P3's annotations on all's summarization.

For validation, we also make a reverse plot in figure 5 which presents the contributions of all's annotations on P3's summarization. We got similar results from this

figure that P2's annotations contributes most to P3's summarization among other four persons. In fact we found most of the annotations of P2 and P3 are similar for most documents. For example, in document "business\_642.txt" whose title is "Stocks Sag; EDS Warnings Whacks IBM Lower", we found that, P2 and both P3 annotated "drooped, Computer makers, Homebuilders, Federal Reserve", which occupies 27% of P2's annotations and 36% of P3's annotations. While in another document, the corresponding values are 36% and 50%. In table 4, we calculate the annotations cosine similarity for any two users, and found that averaged 37% of P3 and P2's annotations are consistent, but 29% for P3 and P5's. This confirms that P2 and P3 should fall into one interest group, while P5 belongs to another.

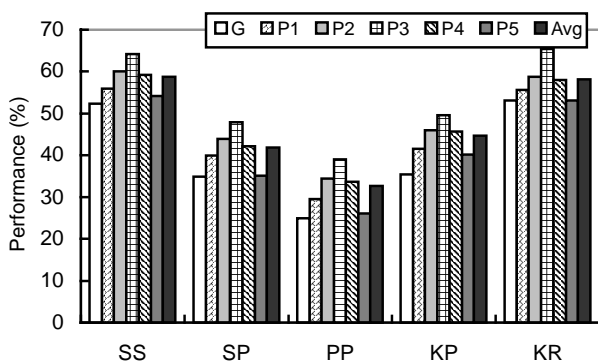


Figure 5. Contributions of all's annotations on P3's summarization.

	P1(%)	P2(%)	P3(%)	P4(%)	P5(%)
P1	100.00	17.90	13.14	10.52	19.83
P2	17.90	100.00	36.87	32.92	29.23
P3	13.14	36.87	100.00	36.02	25.28
P4	10.52	32.92	36.02	100.00	19.40
P5	19.83	29.23	25.28	19.40	100.00

Table 4. Users annotation similarities.

## 6 Conclusion

This paper introduces a new document summarization approach based on users' annotations. Annotations are used to help generate personalized summary for a particular document. Contexts are also considered to make up the incompleteness of annotations. Two types of summaries are produced for each document: generic summary without considering annotations and annotation-based summary. Performance comparisons are made between these two summaries. We show by experiments that annotations are quite useful in improving the personalization of summarization compared to no annotations consideration.

We made extensive evaluations on the relationships between annotations and summaries, and concluded that annotation selection is necessary in order to get a best summarization. We will try to identify how to choose an appropriate subset of annotations that make summarization best, which is challenging due to the variety of users' interests in different documents.

We also did collaborative filtering to find whether one user's annotation is helpful for other's summarization. The answer is positive that summarization performance can be improved based on similar users' annotations. For the next step, we think that "similar users" needs more precise definition to improve the performance of collaborative filtering.

As an extension of collaborative filtering, more work will be done for multi-documents summarization based on annotations of similar ones. They will help user to get a global personal view for a set of documents.

## References

- Edmundson, H.P. 1969. New methods in automatic abstracting. *Journal of the ACM*, 16(2):264-285.
- Eduard Hovy and Chin-Yew Lin. 1997. Automated text summarization in SUMMARIST. In *Proceedings of ACL Workshop on Intelligent Scalable Text Summarization*, Madrid, Spain, July 1997, 18-24.
- Firmin Hand, T. and B. Sundheim. 1998. TIPSTER SUMMAC Summarization Evaluation. *Proceedings of the TIPSTER Text Phase III Workshop*, Washington.
- Gene Golovchinsky, Morgan N. Price, and Bill N. Schilit. 1999. From reading to retrieval: Freeform ink annotations as queries. In *Proceedings of SIGIR'99*, ACM, New York, August 1999, 19-25.
- Gerard Salton and Chris Buckley. 1988. Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 24(5):513-523.
- Gerard Salton and Chris Buckley. 1990. Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science*, 41(4):288-297.
- Julian M. Kupiec, Jan Pedersen, and Francine Chen. 1995. A trainable document summarizer. In *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Washington, July 1995, 68-73.
- Katashi Nagao and Koiti Hasida. 1998. Automatic text summarization based on the Global Document Annotation. In *Proceedings of COLING-ACL*, San Francisco.
- Luhn, H. P. 1958. The automatic creation of literature abstracts. *IBM journal of Research and Development*, 2(2):159-165.
- Mani, Inderjeet and Mark T. Maybury. 1999. *Advances in Automatic Text Summarization*. MIT Press, Cambridge, MA.

- Morgan N. Price, Bill N. Schilit and Gene Golovchinsky. 1998. XLibris: The active reading machine. In Proceedings of CHI '98 Human Factors in Computing Systems, Los Angeles, volume 2 of Demonstrations: Dynamic Documents, 22-23.
- Porter, M.F. 1980. An algorithm for suffix stripping. Program, 14(3):130-137.
- Simone Teufel and Marc Moens. 1997. Sentence extraction as a classification task. In ACL/EACL-97 Workshop on Intelligent Scal-able Text Summarization, Madrid, Spain, July 1997, 58-65.
- Tadashi Nomoto and Yutaka Shinagawa. 2001. A new approach to unsupervised text summarization. In Proceedings of SIGIR 2001, September 2001, New Orleans, LA, 26-34.