Web Content Summarization Using Social Bookmarks: A New Approach for Social Summarization

Jaehui Park
School of Computer Science and Engineering
Seoul National University
Seoul 151-742, Republic of Korea
+82-2-880-5517
jaehui@europa.snu.ac.kr

Tomohiro Fukuhara
RACE (Research into Artifacts, Center for Engineering)
The University of Tokyo, Kashiwa Chiba 277-8568, Japan
+81-4-7136-4275
fukuhara@race.u-tokyo.ac.jp

Ikki Ohmukai
Digital Content and Media Science Research Division, National Institute of Informatics, Tokyo 101-8430, Japan
+81-4-7136-4275
I2k@nii.ac.jp

Hideaki Takeda
Principles of Informatics Research Division, National Institute of Informatics, Tokyo 101-8430, Japan
+81-4-7136-4275
takeda@nii.ac.jp

Sang-goo Lee
School of Computer Science and Engineering
Seoul National University
Seoul 151-742, Republic of Korea
+82-2-880-5517
sglee@europa.snu.ac.kr

ABSTRACT
An increasing number of Web applications are allowing users to play more active roles for enriching the source content. The enriched data can be used for various applications such as text summarization, opinion mining and ontology creation. In this paper, we propose a novel Web content summarization method that creates a text summary by exploiting user feedback (comments and tags) in a social bookmarking service. We had manually analyzed user feedback in several representative social services including del.icio.us, Digg, YouTube, and Amazon.com. We found that (1) user comments in each social service have its own characteristics with respect to summarization, and (2) a tag frequency rank does not necessarily represent its usefulness for summarization. Based on these observations, we conjecture that user feedback in social bookmarking services is more suitable for summarization than other type of social services. We implemented prototype system called SSNote that analyzes tags and user comments in del.icio.us, and extracts summaries. Performance evaluations of the system were conducted by comparing its output summary with manual summaries generated by human evaluators. Experimental results show that our approach highlights the potential benefits of user feedback in social bookmarking services.

Categories and Subject Descriptors
H.3.1 [**Content Analysis and Indexing**]: Abstracting methods; H.3.3 [**Information Search and Retrieval**]: Information filtering, Selection process

General Terms
Experimentation

Keywords
Social summarization, social bookmarking service, user feedback

1. INTRODUCTION
The rapid growth of the World Wide Web has dramatically increased the amount of information. With the expansion of accessible data, it has become more and more difficult for users to find useful information quickly and effectively. Therefore, there is an increasing need to provide summarization in many Web applications. For example, on a search engine such as Google, after a query is issued, a list of URLs is returned accompanied with a snippet for each URL, which gives a brief summary of the target page's content. Many of these summaries are generated by *automatic summarization* tools.

The goal of automatic text summarization is to take an information source, extract content from it, and present the most important content to the user in a condensed form and in a manner sensitive to the user's or the application's needs [1]. Automatic summarization techniques have been studied since the 1950's and are used in many applications. However, there is no effective way to automatically produce high-quality summaries of Web documents similar to a human generated gold-standard. On the other hand, human-constructed summaries such as Web site descriptions in the DMOZ Open Directory Project (ODP)[1] have recently become available. These human-authored summaries give a concise and effective description of Web sites. In this paper, we focus on the human involvement in Web content summarization to capture the main points more effectively.

One feasible data source is a social bookmark. Social bookmarking services provide individual users with an easy way to save links to Web pages that they want to remember or share. Most social bookmarking services encourage users to organize

[1] http://www.dmoz.org

their bookmarks with informal tags or short descriptions. This metadata is produced by human users. As social bookmarking services evolve, user feedback information such as comments and tags are beginning to pile up. The aim of this research is to identify the summarization features of social bookmarking services. We propose a method called “social summarization” for summarizing Web information. By using this method, we can generate text summaries that are of same quality as human-authored summaries.

The rest of this paper is organized as follow. Section 2 shows related work. Section 3 describes our user study on social bookmarking services. Section 4 describes the summarization features of user feedback and the summary generation method. Section 5 presents the results of an experiment. Section 6 presents some discussions, and Section 7 concludes the paper.

2. RELATED WORK

Automatic 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) or task (or tasks) [1]. Much work has been done on automatic text summarization. In [2], Luhn describes a simple, genre-specific technique that uses term frequencies as weights. This was the first study to take a statistical approach to text summarization, and it has had considerable influence on the field. Reference [3] extends Luhn’s work by adding three features (cue phrase, title and heading words, and sentence location). Most commercial summarization tools are based on these basic approaches, and considerable research has been devoted to refinements [4]. Although these approaches have great utility, they depend very much on the particular format or style of writing, such as position in the text or lexical words.

The advent of the World Wide Web has brought with it new challenges for Web page summarization. The structure of a Web page can be used as a guide for summarization. Reference [8] and [11] extract segments of text surrounding hyperlinks between pages. The most accurate sentence is chosen from these segments. This work has shown that summaries that take into account the context information are usually more relevant than those made only from the target document. There has been much work on Web page or Web site summarization exploiting the effect of context in Web [4][5][6][7][9][10].

Our work is related to exploiting extra information in Web applications such as search engines, social bookmarking services and Web directories. Reference [8] analyzes the utility of clickthrough data of a Web search engine, and proposes adapted summarization methods that take advantage of the relationships discovered from the clickthrough data. They adapted Luhn’s text summarization methods and latent semantic analysis [12]. The authors of [4] regard social bookmark tags as user queries to generate query-focused snippets. The fragment ranking and scoring methods are based on their distribution frequency. Reference [6] proposes a summarization method using social relationships in online auctions for summarizing feedback comments. In their investigation, the authors of that study found two types of description to produce a summary. One is a description that appears only in the feedback comment on the target seller, and the other is a description that appears in the feedback comments on sellers other than the target seller but does not appear in the feedback comment on the target seller. All these approaches use information generated by users other than the author of the source content. The above-mentioned research depends on the source itself, they perform poorly for various type of content information such as blog post embedding video clip because they cannot processing the multimedia type content. Our work only focuses on the information created by the user not by the author based on the observation of user behavior pattern trend.

We focus on the behavioral information of users such as annotations and descriptions. Social bookmarking services gather diverse Web content and behavioral information; we call this information and content ‘feedback’ and use it to improve summarization performance. We argue that such feedback can be utilized for *social summarization* of diverse Web content.

3. SOCIAL SUMMARIZATION

3.1 Background

We suggest a new idea for summarization that is inspired by the recent emergence of social services. By encouraging users to submit an opinion on the source content or to participate in a social network, they can express their views on the content in the form of a comment or a review. The point of this approach is based on today’s active user role in the Web 2.0 environment. An increasing number of Web applications are allowing users to play more active roles for enriching the source content. The enriched data can be exploited for summarization. Sites like Amazon.com allow users to submit user reviews and scores on consumer products. YouTube [2] and Flickr [3] gather user commentary on videos and pictures. The related work [4][5][6] call summarization involving social interactions *social summarization*.

Bookmarking is a practice of saving the address of a Web site or a Web page that a user wishes to visit in the future on his/her computer. Social bookmarking, on the other hand, is the practice of saving bookmarks on a Web site or a Web page and tagging them with keywords. This allows many users to describe and organize for one Web content. It is true that more people will take an interest in more valuable information. By permitting users to share their bookmarks, an amount of user feedback can be aggregated in the form of tags, descriptions, comments and so on. In our user study, we found different user behavioral patterns of contributing to Web content in social services such as del.icio.us[4], Digg[5], YouTube, and Amazon.com[6].

3.2 Social Summarization

There is as yet no common definition for social summarization. However, the main point of social summarization is on the users’ attitude toward summarization. Our work focuses on user feedback such as comments or tags. We gather feedback information captured by social bookmarking services such as del.icio.us, extract representative words from the feedback, and score them. A sentence that has many representative words with high scores will be chosen as a summary candidate. These summary candidates may have various aspects; they may be query

[2] http:// www.youtube.com

[3] http:// www.flickr.com

[4] http:// del.icio.us

[5] http://digg.com

[6] http:// www.amazon.com

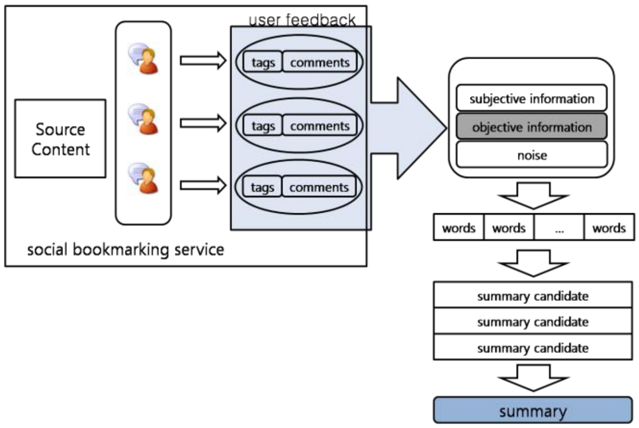

Figure 1. The Concept of Social Summarization

specific or domain specific. Here, we assume that users can choose the best summary from among summary candidates we generate.

However, the involvement of other users who are not the author of the source content makes the summarization work more difficult because of the existence of other user's subjective information. There is much amount of subjective information in user feedback such as an opinion and an impression. In opinion mining field [16], there has been many method for identifying subjective word and objective word. But this kind of technique also removes the author's subjective information in user feedback. To extract the objective information for the source and to filter out the subjective information of other user is one of the issues in social summarization. In this paper, we did not consider this issue, but it is planned to be a future work.

Figure 1 shows the concept of our social summarization approach. To summarize the source content, it is obvious that summarizer should concentrate on the source content itself. There have been various approaches to exploiting key factors in the source such as term frequency, position score and sentence length. However, it is difficult to know what the principal factor is. In social summarization, we entrust users with a task to find these principal factors. We assume that sentences in user feedback which have summarization features are the most important factors in our social summarization method. This conception is similar with the sentiment classification in natural language processing. For example, in online product domain, there has been an assumption that that opinion of users who already buy it is the most important factor for customer's decision making. For our assumption, firstly we should find where the users write a summary for the source content. In following section, we show that users in a social bookmarking service have made a summary for their bookmarks.

3.3 User Study

We investigated user feedback in social services. We chose a representative social bookmarking service: del.icio.us. This is the most popular sites for social bookmarking. Users can put a tag to a bookmark or put a little note on it. YouTube is a video streaming site that allows users to share their videos. Once a user uploads a video, other users can simply rate or submit commentaries on the video. Amazon.com is an online store that allows users to write a review on a commercial product. Additionally, we import investigation results of Digg from our previous work on user feedback [14]. All four sites have a system for sharing user feedback.

Table 1. Statistics of data

Number of evaluator participant	16
Number of source content	1,737
Number of user feedback	117,004
Data Type	Text, Video, Image

First, to ascertain patterns in user feedback, 16 students in a university joined our evaluation test. Table 1 shows the statistics of data. They read user feedback (user notes in del.icio.us, reviews in Amazon.com, commentaries in YouTube and comments in Digg) on a number of articles and Web sites over the course of three months, and 1,737 pieces of content and 117,004 pieces of user feedback were gathered. The user feedback was divided into sentences, and then manually categorized into four groups: *Summary, Additional Information, Opinion, and Noise.*

3.3.1 Data Collection

3.3.1.1 Objective Statement

Objective statements convey information in accordance with the intention of the author. If a user feedback has no judgment or opinion on the source content, we call it objective regardless of whether the content is true or not. There are many forms of objective statement: summary, quotation, comparison with another content, related links, and so on. We divide these into two groups.

Summary is a sentence which explains the gist of the source content. This is a special feature of user feedback patterns in social bookmarking services. Like other social services, a social bookmarking service is a tool for sharing information. But it is basically a tool for remembering information unlikely to be found with other social services. To remember a web page, users make a note or a tag describing the content for their bookmarks. These notes and tags paraphrase the source content and provide a concise representation or extract of the source content. This is the main summarization feature of user feedback in a social bookmarking service. This paraphrases for the source content to provide a concise representation. In summarization field, this is called *abstraction.* On the other hand, it just extracts the parts of text in the source content, which is called *extraction*. These notes and tags are a main summarization factor for our work.

Additional Information is related information from external sources. In brief, it explains facts not appearing in the source content. Additional Information includes quotations, related links or comparisons. For example, "This site is similar with the site A" describes additional information including the site A, and "http://aaa.bbb.com" shows just a link for other page. Although Additional Information is very useful for describing or understanding the source content, this is not used for summarization because of its poor coherence.

3.3.1.2 Subjective Statements

Subjectivity is used to express private states in the context of a text or conversation. "Private state" is a general term for opinions, evaluations, beliefs, emotions, speculations, and so on [13]. Because private states vary from person to person, they sometimes contrast with the intention of the author of the source

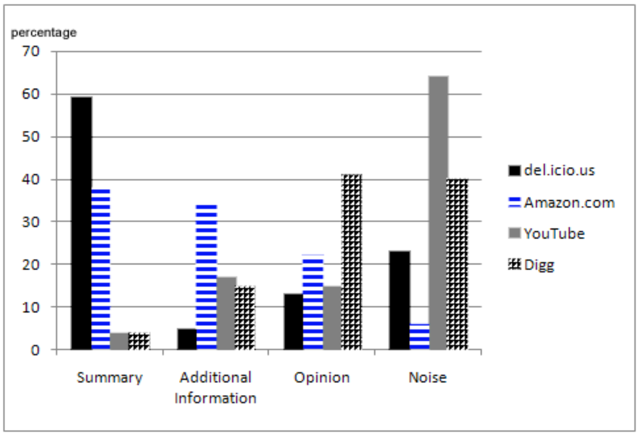

Figure 2. User Feedback Distribution

content. It is destructive to make biased and incoherent summarizations of the source content. However, note that 'subjective' does not mean not true or distrustful. Subjective statements are also important social information and are useful for opinion mining in product reviews or sentiment analysis. However, such analyses are not the focus of our work.

Opinion is a statement that conveys an opinion, evaluation, speculation, assessment, and so on. If there are explicit features of the source content and sentiments associated with them, we categorize such statements as 'opinions'. For example, the sentence, "Interesting introduction.", is classified into Opinion. It describe the introduction part is interesting, but the sentence, "Interesting", describes noting but impression for overall. In this situation, 'introduction' will be the feature word and 'interesting' will be the sentiment word. This information can be used for enriching the source article in social ways.

Noise includes all the irrelevant or low-quality information for summarization including spam, slang, non-English, jokes, and sarcasm. Online conversations will have a certain portion of such noisy feedback. Single-phrase user sentiments such as "Awesome" or "Great job" are considered to be noise. Sometimes, opinions feature sarcasm or are jokes. In a sense, getting the point of such could be complicated even for human evaluators.

In categorizing user feedback, we found that the feedback distribution in the four categories varied according to the social service: del.icio.us, Amazon.com, YouTube, and Digg

3.3.2 Analysis on user feedback among social services

In del.icio.us, over 59% of user notes are of the 'Summary' category. Of the 23.2% classified as 'Noise', many were user notes written in other languages such as Japanese or Spanish. We only analyzed English feedback. The portion for 'Summary' might be even bigger and the portion for 'Noise' smaller if we consider summaries in other languages. This result shows that users of del.icio.us have a tendency to summarize the source content, as we mentioned in Section 3.2. The bookmarks that the user wants to remember are contained in each user page. This page is exposed to other users to share their user notes and tags.

Table 2. Statistics of user feedback

	del.icio.us	Amazon.com	YouTube	Digg
Total amounts	741	130	754	112
Average length	65.72	1001.31	28.17	N/A
Number of feedbacks per content	91.94	11.05	162.12	28.46
Number of tags per content	71.53 (2738)	N/A	N/A	N/A

This social bookmarking process encourages people to make descriptions for the source content rather than discuss the source content. Based on this result, we chose the user feedback in del.icio.us as our experimental data set.

About 64% of YouTube commentaries cannot be processed. Most are jokes or very short expressions of emotion. Users of YouTube express their thoughts swiftly in the form of short commentaries. Most commentaries were categorized as 'Noise'. While some were shrewd criticisms or detailed opinions, the overall results were too noisy for our purpose. Also in Digg, about 40% of user comments are of the 'Noise' category.

The reviews at Amazon.com have a nearly uniform distribution pattern in the categories. 36% were summaries, 35% were additional Information and 21.6% were opinions. They also had very little 'Noise'. Amazon.com includes social information of good quality, which is helpful for customers shopping at the online store. All reviews at Amazon.com are about commercial products. However, there are many biased opinions in reviews, and it becomes very hard to distinguish the 'objective' quality of the reviews [15]. Moreover, book reviews are usually quite lengthy. There is a constraint on feedback length for each site. Users can input 500 characters as a commentary in YouTube, 255 characters in del.icio.us, 350 characters in Digg, but there is no explicit constraint on product review length in Amazon.com.

Table 2 shows that the average numbers for each social service. Tags are also a kind of user feedback, but tag annotations are not allowed in YouTube, Digg and Amazon.com. Therefore we consider tag separately just in del.icio.us. There are over 3 million numbers of tags for 1,737 number of source content, 2738 tags per content. The number of distinct tags per content is 71.53.

Figure 3 shows the plot of top-200 tag frequency in several source contents. It draws an illustration which states that the frequency of any tag is inversely exponential to its rank in the frequency table. The horizontal axis is the rank of a tag in the frequency table. And the vertical axis is the total number of the tag's occurrences.

The most frequently occurring tags are like 'art', 'software' and 'web'. We cannot identify the summarization feature from those kinds of general words. Only 20% of top-9 ranked tags express the gist of the source content, and others are general words. However, about 70% of the tags express the gist of the source

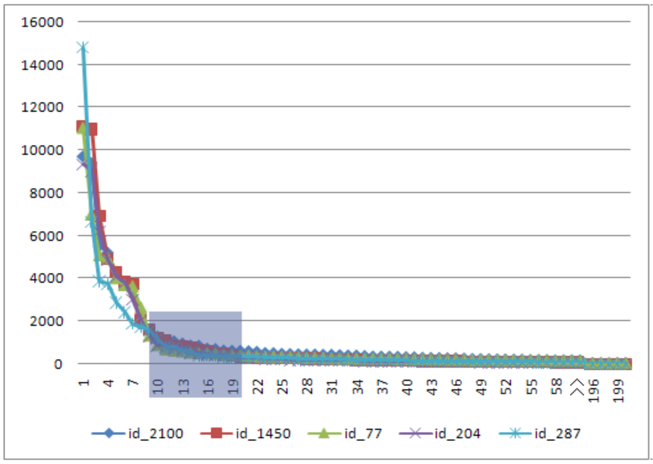

Figure 3. A plot of tag frequency

content within the rank ranging from 10 to 19. It shows certain tendency in occurrences of tags. It implies that a tag frequency rank does not necessarily represent its usefulness for summarization.

4. GENERATING SOCIAL SUMMARIES

From the results of the user study in Section 3.3, we devised a method for summarizing web pages by using user feedback in social bookmarking services. Our social summarization method involves three basic steps: (1) representative word extraction from user feedback, (2) scoring sentences, and (3) summary generation.

4.1 Feature Word Extraction

First, we extract representative words explaining the source content from the user feedback. We use two measures for identifying representative words.

4.1.1 Word Frequency

Frequently mentioned words in overall user feedback are related to the main topic of the source content. For summarization, this relatedness to the main topic is measured based on classical term weighting model (tf-idf) among user feedbacks for the same source content. However, the inverse document frequency factor diminishes the weight of words that explain the gist of the conversation as well as the weight of general words appeared in overall document. This means that classical term weighting model performs poorly in our experiment data set. We suggest a complementary weighting model for the frequency of terms in social conversations:

$$WF(i,j) = \frac{(1+\log(tf_{i,j}))cf_i}{\sqrt{\sum_k^n cf_k^2}} \log\frac{N}{df_i} \qquad (1)$$

$tf_{i,j}$ is the frequency of a word *i* in the user feedback *j*, df_i is the number of user feedbacks that word *i* occurred in, cf_i is the total number of occurrences of word *i* in the overall user feedback, *n* is the total number of words appearing in the source content, N is the total number of user feedbacks, and $WF(i,j)$ is the weight of word *i* in the user feedback *j* based on the term frequency. This model introduces a normalized collection frequency factor into the classical tf.idf model to measure the level of a word's representativeness for the overall user feedback.

4.1.2 Lexicon

We make full use of all words associated with the bookmark. First, the title of the source content is the best summary feature. Second, tags annotating the bookmark are highly related to the source content. Essentially, tags are used for source content categorization. However, some tags express the gist of the source content (see Figure 3). Based on the analysis in Section 3.3.2, we build a set of tags which are included in the portion (10th –19th) over the tag rank list for each associated bookmark. The title and some sets of tags are gathered, and ranked according to the tag's occurrences. Words that appeared in the title are ranked at the top. Let T be the set of tag *i* in the feedback *j*, and let $rank$ the function that returns the rank of tag *i* in the feedback *j*'s tag frequency table.

$$WL(i,j) = \begin{cases} 1+\dfrac{rank(word_{i,j})}{|T+1|} & \text{if } word_{i,j} \in T \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

$WL(i,j)$ is the weight of word *i* in the user feedback *j* based on the occurrences in the set of tags.

After word extraction, we apply stemming and remove stop-words for refinement. The two measures are linearly combined to form Equation (3). In this equation, α is coefficients satisfying $(0 \le \alpha \le 1.0)$

$$Score(i,j) = (1-\alpha)\cdot WF(i,j) + \alpha \cdot WL(i,j) \qquad (3)$$

4.2 Scoring Sentence

Based on the representative words, we extract sentences including those words. We calculate the score of extracted sentence, and normalize it.

4.2.1 Part-of-Speech Tag Sequence

We mentioned the user's pattern of writing a summary for a web page in the previous section (section 3.3.2). It implies that the pattern for constructing a sentence can be captured by a grammar tool for natural language processing (NLP). Based on the user study in Section 3.3, we choose several part-of-speech (POS) tag sequences from user feedback. For example, (NN VBZ RB AT NN) is one of the sequences for a summary sentence like "Time flies like an arrow".

$$P(l) = \frac{1}{1+\sum_i^{\max(m,n)} |c_i - t_i|} \qquad (4)$$

We learn POS tag sequences which are classified into 'summary' class in the user study. Given class C, $<c_1, c_2, \ldots, c_m>$, the selected POS tag sequences *l*, $<t_1, t_2, \ldots, t_n>$, are scored based on Equation (4). $P(l)$ is the score of sequence *l*, which introduces the similarity between two tag sequences.

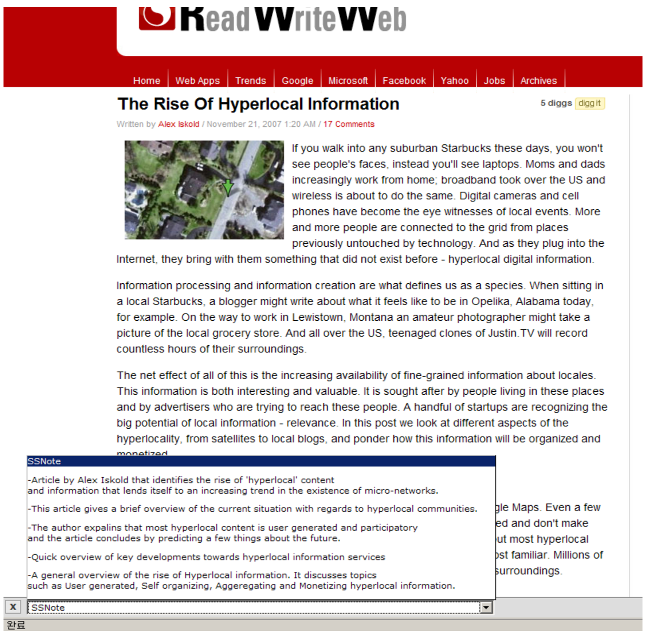

Figure 4. a screen shot image of SSNote

4.2.2 Normalization

The following score based on the normalized sum of word scores (Equation (3)) is used to select the sentence from the user feedback:

$$Score_{sentence}(l) = \frac{P(l)}{N} \sum_{i,l \in j} Score(i, j) \qquad (5)$$

Here, N is the number of words in the sentence. And $Score_{sentence}(l)$ is the score for sentence *l* in the user feedback j.

4.3 Summary Generation

All sentences in the user feedback are ranked by $Score_{sentence}$. To generate one good summary candidate, the compression ratio of summarization should be defined. However, it is difficult to define the compression ratio for a source content including non-text data. We chose the top-*k* sentences from the list of sentences we extracted.

5. EXPERIMENT AND EVALUATION

In order to evaluate the performance of social summarization method, we conducted an experiment. We introduce the system we developed, the experiment data set, and evaluation metrics. Results of our experiment show that our social summarization method works better than the ones without utilizing the user feedback.

5.1 SSNote

We developed prototype system, *SSNote* (Social Summarization Note), which implements our social summarization method. Figure 4 shows a screen shot image of SSNote for a blog page. It is implemented as a browser extension for Mozilla Firefox. SSNote adds a fixed bar at the bottom of the browser window that contains a drop-down list of the social summaries our method generates. This prototype system runs only on web pages or web sites which are bookmarked in del.icio.us.

Table 3. Average summary quality in terms of recall, precision, and F-measure, under ROUGE-1 ad ROUGE-L

	Average	Standard Deviation
ROUGE-1 recall	0.686	0.161
ROUGE-1 precision	0.738	0.082
ROUGE-1 F-measure	0.704	0.115
ROUGE-L recall	0.675	0.165
ROUGE-L precision	0.725	0.090
ROUGE-L F-measure	0.693	0.122

5.2 Data Set

In our experiment, we crawled 1,092 bookmarks from del.icio.us, and randomly sampled 190 bookmarks regardless of their data type, data length or topic of the content. Due to our assumption that there are enough user feedbacks for a bookmark, we only use bookmarks which contain over thirty numbers of user notes, and only written in English.

Two different data sets were used for showing the quality improvement effectively. Whole of the data we sampled is denoted by DAT1. From the randomly sampled bookmarks, we selected bookmarks which were associated with plain text type data such as blog page, denoted by DAT2.

5.3 Evaluation Metrics

In this evaluation, ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [17] is used for measuring summarization quality. ROUGE measures summarization quality by counting overlapping units such as the n-gram, word sequences, and word pairs between the candidate and the gold-standard summaries, a common approach that has been shown to correlate very well with human evaluations [17]. For each bookmark, manually generated summaries are considered as gold-standard summaries. The ROUGE-1 (unigram co-occurrence) metric is highly effective for single document summarization evaluation of short summaries. The ROUGE-L (longest common subsequence) is used because it counts only in-sequence co-occurrences [17].

5.4 Experimental Results

5.4.1 A Comparison of Summary Quality

We will show the overall summary quality by comparing the summary generated by SSNote to the gold-standard using the ROUGE measures.

The results are presented in Table 3. Baseline results are presented in parentheses. We evaluate the quality of the summaries generated by our summarization method to comparable summaries generated by state-of-the-art summarization system, MEAD [18]. For each summary we also generate a comparable MEAD summary of the same length for the purpose of comparison.

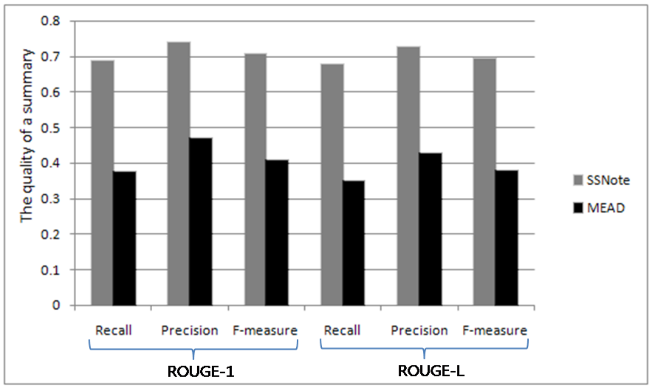

Figure 5. Overall summary quality for SSNote and MEAD in DAT1

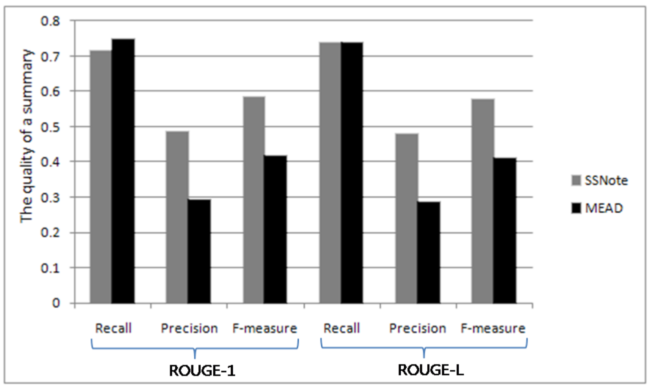

Figure 6. Overall summary quality for SSNote and MEAD in DAT2

Overall summary quality for each technique (SSNote, MEAD) is presented in Figure 5. Experiment results show a great benefit to SSNote across all metrics. This implies that the social summary and the human generated summary are very similar in comparison with the output of other summarization system. Based on this result, our approach highlights the potential benefits of user feedbacks in social bookmarking services. These improvements are consistent across all different data types. On the other hand, MEAD system is a text summarizer. Therefore it doesn't perform well on the data of various types such as video. Therefore we perform same evaluation again on the text data (DAT2).

The results are presented in Figure 6. It is shown that SSNote achieves a relative improvement in its precision and F-measure scores over MEAD by between 65% and 67%. But its recall is almost same or 4% lower than MEAD. Based on this result, our approach significantly out-performs MEAD across the relevancy to the source content in text summarization. And it has at least same quality for the sensitivity of MEAD.

5.4.2 User comment and Tag

So far we have demonstrated the improvement of our approach. In this section, we will consider the influence of the two measures of our social summarization method on summary quality.

In SSNote, user notes and tags are involved to extract representative words. α is the ratio of tag use to overall feedback use. We look at the changes of summary quality by changing the

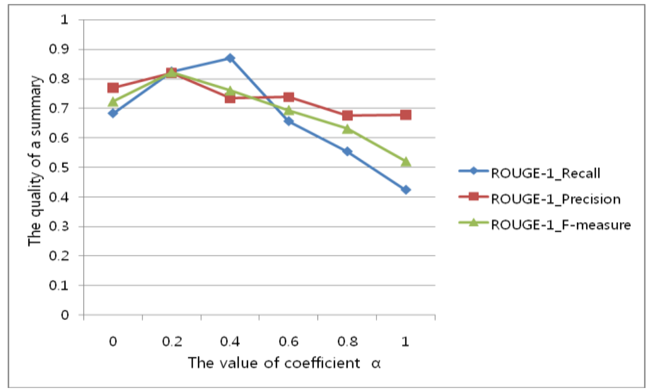

Figure 7. Summary quality according to coefficient α

value of α . Figure 7 shows the results. This figure shows that our system performs best when α is 0.2. However, as α increase we see a decrease in the summary quality on the whole. This implies that user comments are more suitable for creating a summary than tags. But combining the two measures will benefit when a certain coefficient value is determined.

6. DISCUSSION

Our experimental results clearly highlight the potential benefits of social summarization closely related to a human-generated gold standard. It shows an intuitive and simple form of summarization involving humans.

So far, we have found that our summarization method works on the assumption that there exists certain amount of user feedback. This means there is a performance problem when there are few user feedbacks for a web page. It's not because of the technique but because of the social information availability. In our experiment, we fixed the low bound of the number of user feedback. This assumption is the great issue for our social summarization method.

The patterns of writing user feedback vary among people. It is difficult to analyze the grammatical structure of a good summary with shallow parsing or part-of-speech tagging. In this paper, we focused more on term frequency measures than on grammatical factors. In the future, we will try to develop a more specific feature extraction model for learning the writing user feedback pattern.

Our method results in several summary candidates. From these summaries, a user can choose the best summary for his/her context. This method still needs an automatic way to complete the whole process of social summarization, and this will be treated in a future work.

7. CONCLUSION

This paper presented a novel Web content summarization method that creates a text summary by exploiting user feedback (comments and tags) in a social bookmarking service Our user study found that (1) user comments in each social service have its own characteristics with respect to summarization, and (2) a tag frequency rank does not necessarily represent its usefulness for summarization. Based on these observations, we conjecture that user feedback data in social bookmarking services is more suitable for summarization than other type of social services. Our social summarization method first extracts feature words and then scores the sentence which contains the feature words. We implemented prototype system called SSNote that analyzes tags and user comments in del.icio.us for the evaluation of our social summarization method. Performance evaluations of the system were conducted by comparing its output summary with manual summaries generated by human evaluators. Experimental results show that our approach highlights the potential benefits of user feedback in social bookmarking services.

8. ACKNOWLEDGMENTS

This research was supported by the Ministry of Knowledge Economy, Korea, under the Information Technology Research Center support program supervised by the Institute of Information Technology Advancement. (grant number IITA-2008-C1090-0801-0031)

9. REFERENCES

[1] Mani, I. and Maybury, M. T. 1999. Advances in Automatic Summarization. The MIT Press

[2] Luhn, H. P. 1958. The Automatic Creation of Literature Abstracts. IBM J. Res. Develop. 2. 159-165

[3] Edmundson, H. P. 1969. New Methods in Automatic Extracting. J. ACM 16, 2 (Apr. 1969), 264-285. DOI= http://doi.acm.org/10.1145/321510.321519

[4] Boydell, O. and Smyth, B. 2007. From social bookmarking to social summarization: an experiment in community-based summary generation. In Proceedings of the 12th International Conference on Intelligent User Interfaces (Honolulu, Hawaii, USA, January 28 - 31, 2007). IUI '07. ACM, New York, NY, 42-51. DOI= http://doi.acm.org/10.1145/1216295.1216311

[5] Hu, M., Sun, A., and Lim, E. 2007. Comments-oriented blog summarization by sentence extraction. In Proceedings of the Sixteenth ACM Conference on Conference on information and Knowledge Management (Lisbon, Portugal, November 06 - 10, 2007). CIKM '07. ACM, New York, NY, 901-904. DOI= http://doi.acm.org/10.1145/1321440.1321571

[6] Hijikata, Y., Ohno, H., Kusumura, Y., and Nishida, S. 2006. Social summarization of text feedback for online auctions and interactive presentation of the summary. In Proceedings of the 11th international Conference on intelligent User interfaces (Sydney, Australia, January 29 - February 01, 2006). IUI '06. ACM, New York, NY, 242-249. DOI= http://doi.acm.org/10.1145/1111449.1111500

[7] Shen, D., Chen, Z., Yang, Q., Zeng, H., Zhang, B., Lu, Y., and Ma, W. 2004. Web-page classification through summarization. In Proceedings of the 27th Annual international ACM SIGIR Conference on Research and Development in information Retrieval (Sheffield, United Kingdom, July 25 - 29, 2004). SIGIR '04. ACM, New York, NY, 242-249. DOI= http://doi.acm.org/10.1145/1008992.1009035

[8] Sun, J., Shen, D., Zeng, H., Yang, Q., Lu, Y., and Chen, Z. 2005. Web-page summarization using clickthrough data. In Proceedings of the 28th Annual international ACM SIGIR Conference on Research and Development in information Retrieval (Salvador, Brazil, August 15 - 19, 2005). SIGIR '05. ACM, New York, NY, 194-201. DOI= http://doi.acm.org/10.1145/1076034.1076070

[9] Zhang, Y.Zincir-Heywood N. and Milios, E. 2003. Summarizing Web Sites Automatically. Lecture Notes in Computer Science 2620. Springer, Berlin, 185-199.

[10] Amitay, E. and Paris, C. 2000. Automatically Summarising Web Sites: is there a way around it? In Proceedings of the Ninth international Conference on information and Knowledge Management (McLean, Virginia, United States, November 06 - 11, 2000). CIKM '00. ACM, New York, NY, 173-179. DOI= http://doi.acm.org/10.1145/354756.354816

[11] Delort, J., Bouchon-Meunier, B., and Rifqi, M. 2003. Enhanced web document summarization using hyperlinks. In Proceedings of the Fourteenth ACM Conference on Hypertext and Hypermedia (Nottingham, UK, August 26 - 30, 2003). HYPERTEXT '03. ACM, New York, NY, 208-215. DOI= http://doi.acm.org/10.1145/900051.900097

[12] Gong, Y. and Liu, X. 2001. Generic text summarization using relevance measure and latent semantic analysis. In Proceedings of the 24th Annual international ACM SIGIR Conference on Research and Development in information Retrieval (New Orleans, Louisiana, United States). SIGIR '01. ACM, New York, NY, 19-25. DOI= http://doi.acm.org/10.1145/383952.383955

[13] Wiebe, J., Wilson, T., Bruce, R., Bell, M., and Martin, M. 2004. Learning Subjective Language. Comput. Linguist. 30, 3 (Sep. 2004), 277-308. DOI= http://dx.doi.org/10.1162/0891201041850885

[14] Park, J., Fukuhara, T., Ohmukai, I., and Takeda, H. 2008. Web Content Summarization Using Social Bookmarking Service. NII Technical Report. National Institute of Informatics of Japan.

[15] David, S and Pinch, T. J. 2005. Six Degrees of Reputation: The use and abuse of online reviews and recommendation systems. First Monday 11, 3. (Nov. 2005)

[16] Dave, K., Lawrence, S., and Pennock, D. M. 2003. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international Conference on World Wide Web* (Budapest, Hungary, May 20 - 24, 2003). WWW '03. ACM, New York, NY, 519-528. DOI= http://doi.acm.org/10.1145/775152.775226

[17] Lin, C. and Hovy, E. 2003. Automatic evaluation of summaries using N-gram co-occurrence statistics. In *Proceedings of the 2003 Conference of the North American Chapter of the Association For Computational Linguistics on Human Language Technology - Volume 1* (Edmonton, Canada, May 27 - June 01, 2003). North American Chapter Of The Association For Computational Linguistics. Association for Computational Linguistics, Morristown, NJ, 71-78. DOI= http://dx.doi.org/10.3115/1073445.1073465

[18] Radev, D., Allison, T., Blair-Goldensohn, S., Blitzer, J., Celebi, A., Dimitrov, S., Drabek, E., Hakim, A., Lam, W., Liu, D., Otterbacher, J., Qi, H., Saggion, H., Teufel, S., Topper, M., Winkel, A., and Zhang, Zhu. 2004. MEAD - a platform for multidocument multilingual text summarization. In Proceedings of the 4th international Conference on Language Resources and Evaluation (Lisbon, Portugal, May 24-30, 2004). LREC’04. ELDA, Paris 55-57