Computing and Informatics, Vol. 34, 2015, 779–794

EFFICIENT KEYWORD EXTRACTION AND TEXT SUMMARIZATION FOR READING ARTICLES ON SMART PHONE

Hyoungil Jeong

Department of Computer Science and Engineering, Sogang University
1 Sinsu-dong, Mapo-gu, 121-742
Seoul, Republic of Korea
e-mail: hijeong@gmail.com

Youngjoong Ko[*]

Computer Engineering, Dong-A University
840 Hadan 2-dong, Saha-gu, 604-714
Busan, Republic of Korea
e-mail: youngjoong.ko@gmail.com

Jungyun Seo

Department of Computer Science and Engineering, Sogang University
1 Sinsu-dong, Mapo-gu, 121-742
Seoul, Republic of Korea
e-mail: seojy@sogang.ac.kr

Abstract. These days, we can connect to the internet from almost anywhere, allowing us to access web content, including newspapers, magazines, blogs and websites, using mobile devices such as a smart phone. However, people sometimes struggle to read and use the contents due to the nature of these devices such as a small display, low display resolution and limited computing resources (low CPU speed and little memory). This paper aims to develop a convenient interface that provides keyword

[*] corresponding author

extraction, summary generation and search engine to users. We apply the proposed summarization method to Korean and English news articles and evaluate it using several experiments on single and multiple news article test collections and user-receptiveness tests. Since the proposed method shows a good performance on these experiments and tests, we think that this interface can help users more efficiently to read the news articles on various mobile devices.

Keywords: Single and multiple news article summarization, keyword extraction, query expansion, mobile devices

Mathematics Subject Classification 2010: 68-U35, 68-T50

1 INTRODUCTION

As a smart phone becomes more and more popular, access to the World Wide Web from smart phone applications is an exciting and promising addition to our web experience. However, users usually struggle when using a smart phone to access the web, because they have no choice but to use the small screen. Moreover, because the amount of content within each page is too large to be adequately displayed, converting standard web pages shown on a smart phone is a challenge [1].

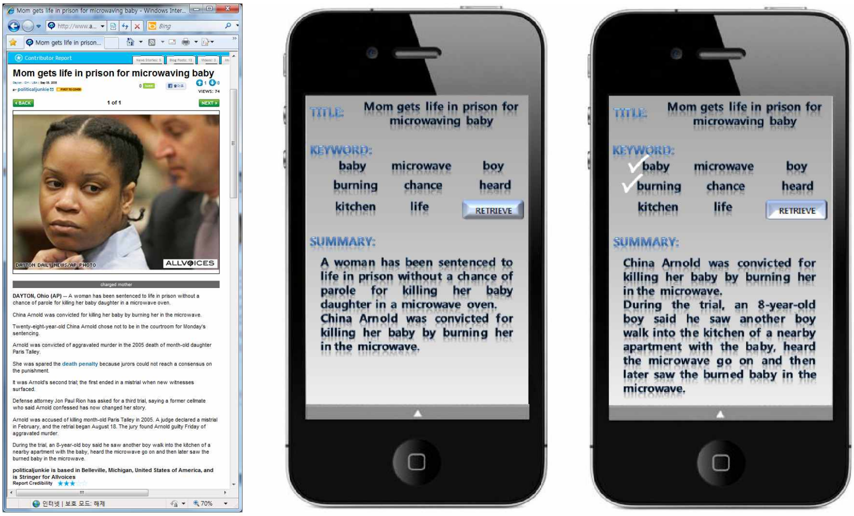
Since the web pages that are most frequently read are news articles, they are selected as a main application domain for our research. Instead of requiring that people read all the content in a news article, we propose that providing summaries of the article is one way to save the time. However, most news articles do not include a summary due to the high cost of manually summarizing them. Thus, we can apply automatic keyword extraction and text summarization techniques to develop an efficient interface for a smart phone. This will reduce the complexity and length of the news articles from the web, while retaining the essential qualities of the original news articles. Furthermore, users often search for news articles using a search engine on a smart phone by querying a particular topic. Since these searched articles can be considered to be multiple news articles for a topic, the multiple news article summarization can be useful when using a smart phone.

In this paper, we propose an efficient interface for an easy reading news articles and searching for information from them on a smart phone. The proposed system can provide keywords, a summary for single and multiple articles, and search for the user information. Since the system has limited CPU resources and memory, we use only statistical methods for extracting keywords and summary sentences instead of machine learning methods or linguistic approaches.

The rest of this paper is organized as follows. Section 2 shows our proposed interface that produce keywords and summary. In Section 3, we discuss related works. Section 4 explains the implemented system, which consists of three modules: keyword extraction, summary sentence extraction and a search engine. Section 5 discusses the experimental results and Section 6 contains a summary and ideas for the future work.

2 INTERFACE

An example of the interface for showing keywords and summary results is illustrated in Figure 1. This interface is for a single news article. You can see three different frames in the interface: the TITLE frame, the KEYWORD frame and the SUMMARY frame.



Mom gets life in prison for microwaving baby

by politicaljunkie

DAYTON, Ohio (AP) -- A woman has been sentenced to life in prison without a chance of parole for killing her baby daughter in a microwave oven.

China Arnold was convicted for killing her baby by burning her in the microwave.

Twenty-eight-year-old China Arnold chose not to be in the courtroom for Monday's sentencing.

Arnold was convicted of aggravated murder in the 2005 death of month-old daughter Paris Talley.

She was spared the death penalty because jurors could not reach a consensus on the punishment.

It was Arnold's second trial; the first ended in a mistrial when new witnesses surfaced.

Defense attorney Jon Paul Rion has asked for a third trial, saying a former cellmate who said Arnold confessed has now changed her story.

Arnold was accused of killing month-old Paris Talley in 2005. A judge declared a mistrial in February, and the retrial began August 18. The jury found Arnold guilty Friday of aggravated murder.

During the trial, an 8-year-old boy said he saw another boy walk into the kitchen of a nearby apartment with the baby, heard the microwave go on and then later saw the burned baby in the microwave.

politicaljunkie is based in Belleville, Michigan, United States of America, and is Stringer for Allvoices
Report Credibility ★★★☆☆

Mom gets life in prison for microwaving baby

A woman has been sentenced to life in prison without a chance of parole for killing her baby daughter in a microwave oven. China Arnold was convicted for killing her baby by burning her in the microwave.

Mom gets life in prison for microwaving baby

China Arnold was convicted for killing her baby by burning her in the microwave.
During the trial, an 8-year-old boy said he saw another boy walk into the kitchen of a nearby apartment with the baby, heard the microwave go on and then later saw the burned baby in the microwave.

a) b) c)

Figure 1. Example of a single news article summarization results: a) an original news article, b) the title, keywords and summary results and c) a summary by keyword retrieval

TITLE frame: the title of the original newspape.

KEYWORD frame: the keywords that are extracted as significant terms by our statistical method are shown in this frame. Checkboxes are provided to generate queries in the search engine. If the user selects the checkboxes of interesting keywords and then presses the RETRIEVE button, the system retrieves and presents those sentences that include the keywords. The sentences are extracted using a query that is composed of the selected keywords. This retrieval function is very useful when users want to see the sentences but do not have access to a keyboard.

SUMMARY frame: summary sentences are shown in this frame. The summary sentences are extracted using our proposed summarization method.

We design an efficient interface to solve some readability problems when reading news articles through a smart phone and developed core technologies to implement the interface such as keyword extraction and text summarization. For evaluation, we conducted experiments on the test sets for single and multiple news articles and the user receptiveness tests for summarization. As a result, we verified the efficiency and effectiveness of the proposed method in our evaluation.

3 RELATED WORK

A number of researchers have been studying text summarizing methods for small display devices with limited resources.

Buyukkokten et al. proposed a text summarization method for browsing the web on handheld devices such as personal digital assistants (PDA) [2]. They suggested a system to break a web page into text units that can be hidden, partially displayed, made fully visible, or summarized by using a simple statistical calculation. They named the text units semantic textual units (STUs) and arranged them appropriately on the display, helping to make the content more readable. Rahman et al. discussed the issue of a small form factor view for electronic devices from the perspective of web browsing [3]. They proposed an approach to automatically summarize and transform web documents into meaningful, readable and browsable content. Jones et al. carried out experiments to explore the actual effects of small displays on users abilities to interact with web pages originally designed for conventional and large screen displays [4]. Sweeney and Crestani investigated the effects of the length of the summary as a function of screen size, in which query-biased summaries are used to present retrieval results [5]. They reported a user study aimed at exploring whether there is an optimal summary size for three types of devices, given their different screen sizes.

Methods of efficiently summarizing texts have also been studied by many researchers. Berger and Mittal proposed a web page summarization system that can generate coherent summaries that are not excerpts from the original document [6]. Harabagiu and Lăcătuşu proposed an information extraction based multi document summarization procedure that incrementally adds information [7]. They have shown that it is possible to obtain a good quality multi-document summaries by using extraction templates populated by a high performance information extraction system. Matsuo and Ishizuka developed an algorithm that extracts keywords from a single document [8]. Their algorithm can extract keywords without requiring the use of a corpus. They stated that their method has a higher performance than the term frequency-inverse document frequency (TF-IDF) and that their method is useful in many applications, especially for domain-independent keyword extraction. Svore et al. proposed an automatic summarization method based on neural nets, called NetSum [9]. They extracted a set of features from each sentence that help to identify its importance in the document and then applied novel features based on news search query logs and Wikipedia entities using the RankNet learning algorithm. Ko and

Seo proposed an effective method for extracting salient sentences using contextual information and statistical approaches for text summarization [10]. They combined two consecutive sentences into a bi-gram pseudo sentence so that contextual information is applied to statistical sentence-extraction techniques. Li et al. proposed a document summarization approach, named ontology enriched multi document summarization, for utilizing background knowledge to improve summarization results [11]. Their proposed system can better capture the semantic relevance between the query and the sentences and thus leads to better summarization results by using the domain-related ontology.

4 THE PROPOSED SYSTEM

The proposed system consists of three parts: keyword extraction, summary sentence extraction and a search engine. Each part is described in the following sections in detail. The overview of our system is illustrated in Figure 2.

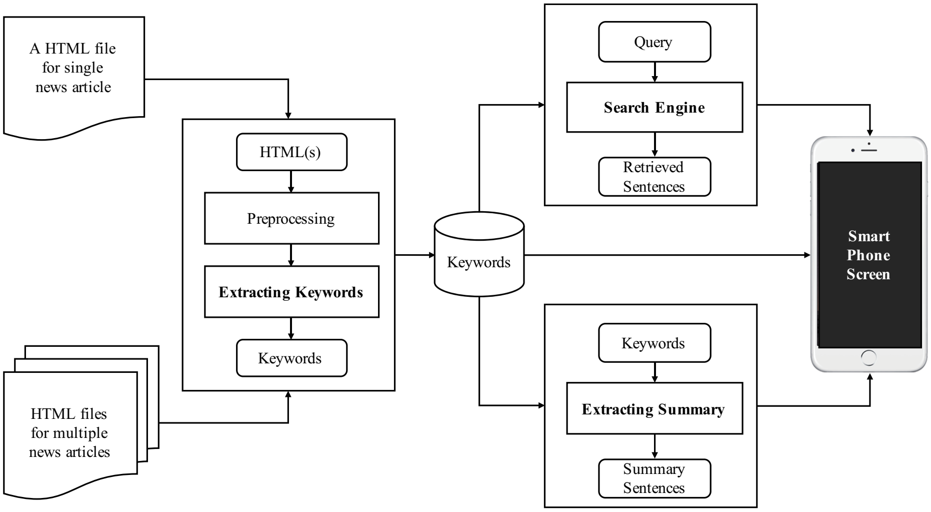


Figure 2. Overview of the proposed system

4.1 Extracting Keywords

In this section, we explain how to extract keywords using the proposed method. This process consists of three steps. The first step is for a preprocessing task in which an HTML document is converted to a matrix form. The second step selects relevant sentences and the third step estimates the importance of the terms.

4.1.1 Preprocessing

The proposed system takes an HTML file containing a news article as an input. This file needs to represent vectors from the article [12]. First, it extracts the title and the body of the news article from the HTML file. Then the news article is separated into sentences and they are analyzed into morphemes by a POS tagger [13]. Finally, all the sentences of the news article are converted into a matrix, D, in which each row of the matrix corresponds to a sentence and a column corresponds frequencies of a noun term in Korean news article. If a news article is written in English, a column of matrix, D, corresponds frequencies of noun or verb terms.

4.1.2 Selecting the Relevant Sentences

Pseudo relevance feedback takes the results that are initially returned from a query and uses information about the relevance of the results to perform a new query [14]. We apply this basic idea of the pseudo relevance feedback to our system to effectively extract keywords and summary sentences. In our system, the noun terms from a title are regarded as initial query terms. Sentences that include the initial query terms are regarded as relevant sentences.

4.1.3 Estimating the Importance of Keyword Candidates Using Statistical Relevance Weighting

All the noun terms of the relevant sentences become keyword candidates. Some keyword candidates with high importance are selected as keywords. The importance of each keyword candidate is estimated by a relevance weighting function depending on the distribution of the keyword candidate in relevant and non-relevant sentences. This is based on the *Binary Independence Model (BIM)* [15] proposed by Robertson and Jones.

$$\begin{aligned} W_{kc} &= \log \frac{p_{kc}(1-q_{kc})}{(1-p_{kc})q_{kc}} \\ &= \log \frac{\frac{r_{kc}+0.5}{R}\left(1-\frac{s_{kc}+0.5}{S}\right)}{\left(1-\frac{r_{kc}+0.5}{R}\right)\frac{s_{kc}+0.5}{S}} \\ &= \log \frac{(r_{kc}+0.5)(S-s_{kc}+0.5)}{(R-r_{kc}+0.5)(s_{kc}+0.5)}, \qquad (1) \end{aligned}$$

$$\begin{aligned} p_{kc} &= \frac{\#\text{ of sentences that include } kc \text{ in } R}{R}, \\ q_{kc} &= \frac{\#\text{ of sentences that include } kc \text{ in } S}{S}, \end{aligned}$$

where W_{kc} is the relevance weight of a keyword candidate kc; p_{kc} and q_{kc} are the probabilities that kc appears in relevant and non-relevant sentences, respectively; R and S are the number of relevant sentences (include any words of the title) and non-relevant sentences (include no words of the title) in an article, respectively; and r_{kc} and s_{kc} are the number of relevant and non-relevant sentences that include kc, respectively.

After all, the keyword candidates are sorted according to their relevance weights, the top m keyword candidates are selected keywords. These keywords are used in the KEYWORD frame of Figure 1. and in calculating sentence weights for summary generation. In the KEYWORD frame, the number of keywords can be determined according to the GUI design of a smart phone. On the other hand, the number of keywords in the sentence weight calculation is set to 5 by our experiments.

4.1.4 Extracting Summary Sentences

In order to select summary sentences, we estimate the importance score of each sentence by using the relevance weights of keywords and the position information of the sentence within an article [16, 17]. We first calculate the sum of the keyword weights to apply the keyword weights estimated by Equation (1) to the importance of sentences in Equation (2). We consider sentences that include keywords with a high relevance weight to be more important.

$$\text{KeywordScore}(S_i) = \sum_{k_j \in S_i} W_{k_j}. \tag{2}$$

In Equation (2), W_{k_j} denotes the relevance weight of j^{th} keyword k_j, included in i^{th} sentence S_i, and $\text{KeywordScore}(S_i)$ denotes the importance score of S_i that is calculated by the sum of relevance weights of keywords in S_i.

The leading sentences at the beginning of news articles are considered to be more important sentences [18]. In order to apply this idea to the proposed method, we add position information to the final equation for estimating the importance score of each sentence by linear combination as follows:

$$\text{score}(S_i) = \alpha \left(\frac{\text{KeywordScore}(S_i)}{\text{KeywordScoreMax}} \right) + (1 - \alpha) \left(1 - \frac{i-1}{N} \right), \tag{3}$$

where KeywordScoreMax denotes the highest importance score of sentence within a news article, N is the total number of sentences in the news article, i denotes the position information of S_i and α is a parameter for linear combination, which is set to 0.4 according to the results of our experiments.

As a result, we select the top n sentences with high importance scores as a summary. For example, if you want a $30\,\%$ summarization then $n = 0.3 \times N$.

4.2 Applying the Proposed Method to Multiple News Article Summarization

In order to apply our proposed method to multiple news article summarization, we need to make some changes to the proposed summary extraction method. Since each news article in a topic has its own title, we can first select relevant sentences from each news article by using its title. Then all these relevant sentences from each news article are merged into a set of relevant sentences for the topic and all the other sentences are considered as non-relevant sentences for the topic. Using these relevant and non-relevant sentence sets of the topic, we estimate the relevance weights of keyword candidates and choose keywords for the topic; the number of keywords is here set to 8 by our experiments. Then we calculate sentence importance scores by using Equation (4) just like Equation (2). In addition, the position information is also calculated in each news article; the leading sentences of each news article are considered to be more important. The final importance scores for all the sentences in a multiple news article set with the same topic are calculated by the following Equations (4) and (5).

$$\mathrm{KeywordScore}(S_{i_d}) = \sum_{k_j \in S_{i_d}} W_{k_j}, \tag{4}$$

where S_{i_d} denotes i^{th} positioned sentence in a news article, d, in a topic.

$$score(S_{i_d}) = \alpha \left(\frac{\mathrm{KeywordScore}(S_{i_d})}{\mathrm{KeywordScoreMax}_d} \right) + (1 - \alpha) \left(1 - \frac{i_d - 1}{N_d} \right), \tag{5}$$

where $\mathrm{KeywordScoreMax}_d$ denotes the highest importance score of sentence in d, i_d denotes the position information of S_{i_d} and N_d is the total number of sentences in d.

The sentences with a high value of $\mathrm{score}(S_{i_d})$ in a topic are selected as a summary of the topic.

4.3 The Search Engine

We provide a simple search engine to users to enable them to easily find interesting topics (sentences) from news articles. Since we assume that users cannot use the peripherals of a desktop computer such as a keyboard, we exploit the extracted keywords as the input query of our search engine. That is, users can see keywords and a summary on the screen of the smart phone, click some checkboxes for interesting keywords and search for sentences related to the keyword. Our search engine is implemented by using the *TF-IDF* scheme, the *cosine-similarity* measure and the *inverted indexing* method, in order to retrieve and sort sentences. We think that this can offer users a very efficient way of finding interesting sentences in single or multiple news articles on a smart phone.

5 EMPIRICAL EVALUATION

We evaluated the proposed system using two different evaluation methods. The first is *performance evaluation* of Korean and English newspapers and the second is *user receptiveness evaluation* by human testers.

5.1 Performance Evaluation

In this section, we first evaluate the performance of the proposed system for single and multiple news article summarization.

5.1.1 Data Sets and Experimental Settings

For Korean single news article summarization, we used the KORDIC (KOrea Research and Development Information Center) test collection [19, 20] that is well-known in the field of testing Korean single news article summarization. This data is composed of 841 Korean news articles. Each news article has a title, contents and a 30 % summary. The average number of sentences in a news article is 16.37 and the average content words in a sentence are 11.97. We used 252 articles as a validation set in order to decide the parameters such as α.

Item	Statistics
The # of Articles	841
Average of Sentences in an Article	16.37
Average of Content Words in a Sentence	11.97

Table 1. Statistics of the Korean single news article summarization data set

In the case of multiple Korean news article summarization, we used a test collection constructed by Ko's and Seo's research [10, 20]. This data set has 5 topics of 55 articles and 949 sentences and a hand-constructed 10 % summary for each topic. The composition of the Korean multiple news article summarization data set is shown in Table 2.

Topic	The # of Articles	The # of Sentences	Topic Title
1	8	109	A Korean Actress' Nude Scandal
2	15	190	Dr. Hwang, Woo-suk: The 'Stem Cell' Man
3	14	324	Korean Movies
4	11	166	Spain Terror
5	7	160	North Korea's Nuclear
total	55	949	–

Table 2. Composition of the Korean multiple news article summarization data set

Unfortunately, we cannot find an appropriate English single and multiple news article summarization data set which is similar to above Korean data sets. Thus we constructed an English news summarization test collection [20] in the same manner as the Korean news summarization test collections above. This test data consists of 100 articles with 7 topics from The New York Times (NYT) [21].

For single English news article summarization, we annotated 30 % summary in each article. The average number of sentences in a news article is 17.69 and the average content words in a sentence is 26.28.

Item	Statistics
The # of Articles	100
Average of Sentences in an Article	17.69
Average of Content Words in a Sentence	26.28

Table 3. Statistics of the English single news article summarization data set

In the case of multiple English news article summarization, we annotated 10 % summary in each topic. The composition of this new data set is shown in Table 4.

Topic	The # of Articles	The # of Sentences	Topic Title
1	15	238	1 year after Nuclear Disaster in Japan
2	14	285	Apple Siri
3	14	263	Kate Middleton Marriage
4	14	216	New iPad
5	14	245	North Korean Nuclear Program
6	14	244	Occupy Movement (Occupy Wall Street)
7	15	278	Syrian Opposition
total	100	1769	-

Table 4. Composition of the English news article summarization data set

There are many evaluation measures for text summarization [22]. We used meat-and-potatoes measures. The performance of our system is measured by the following equations as *precision*, *recall* and *F1-measure*.

$$\textit{precision} = \frac{\#\text{ of correct summary sentences in created summary sentences}}{\#\text{ of created summary sentences}}, \tag{6}$$

$$\textit{recall} = \frac{\#\text{ of correct summary sentences in created summary sentences}}{\#\text{ of correct summary sentences}}, \tag{7}$$

$$\textit{F1-measure} = \frac{2 \times \textit{precision} \times \textit{recall}}{\textit{precision} + \textit{recall}}. \tag{8}$$

5.1.2 Other Methods for Comparison

We used four other statistical methods for comparing single news article summarization results, including *Title method* [18], *Location method* [18], *Frequency method* [23] and *Aggregation method* [24]. These methods are well-known as fast single document summarization methods. Furthermore, we used three other statistical methods for comparing multiple news article summarization results, including *Contextual Information* [10], *Title and Location method* [18] and *Maximum Marginal Relevance (MMR)* [25]. These methods are summarized in Table 5.

Method	Description	Comparison
Title method	Sentences that have high cosine-similarity with the title are extracted as a summary.	Single news article
Location method	The beginning sentences of article are extracted as a summary.	Single news article
Title and location method	A linear combination of the title method and the location method	Multiple news article
Frequency method	Sentences that have high sum of *TF-IDF* values of terms are extracted as a summary.	Single news article
Aggregation method	Sentences that have high sum of similarity to other sentences are extracted as a summary.	Single news article
Contextual method	Sentences that have maximum contextual information are extracted as a summary.	Multiple news article
MMR	Sentences that have minimum redundancy are extracted as a summary.	Multiple news article

Table 5. Statistical methods for comparing news article summarization methods

5.1.3 Experimental Results

As can be seen from Table 6, the proposed system achieved the best performance among statistical methods used in experiments for single Korean and English news article summarization in the 30 % summary.

Methods	Korean	English
Proposed System	0.511	0.623
Title Method	0.482	0.536
Location Method	0.494	0.612
Frequency Method	0.378	0.519
Aggregation Method	0.415	0.614

Table 6. Experimental results for single news article summarization (F-1 Measure)

In Table 7, the proposed system led to the best performance among the multiple Korean and English news article summarization methods on the 10 % summary.

Methods	Korean	English
Proposed System	0.540	0.423
Contextual Information Method	0.516	0.401
Title and Location Method	0.479	0.392
MMR (λ=0.3)	0.482	0.323
MMR (λ=0.7)	0.483	0.394

Table 7. Experimental results for multiple news article summarization (F-1 Measure)

Even though the *MMR method* has been generally known as a superior summarization method on multiple text summarization due to its consideration of the information redundancy, it showed a poor performance in our experiments. The reasons why we obtained those results are that news articles rarely include the same sentences, and many sentences from news articles with the same topic are written by using the same topic words such as proper noun.

5.2 User Receptiveness Evaluation

Because the quality of summaries is dependent on users, it is very difficult to evaluate them using only quantitative analysis. Therefore, we also conducted user receptiveness evaluation tests on 20 testers and analyzed the results. Our findings appear in Tables 8 and 9. Test 1 is an evaluation for the summary receptiveness test shown in Table 8 and Test 2 is the summary method test shown in Table 9.

In the summary receptiveness test as Test 1, we estimated the extent to which a summary reflects the contents of the original news article. We showed 20 randomly selected pairs of an original news article and a summary to each tester. Each original news article had 15 ± 2 sentences. We used a 7-point measure in all the user receptiveness tests; 7 is the best score and 1 is the worst score. As can be seen from Table 8, we achieved high scores fairly in Test 1.

	Question	Korean	English
Q1	(show an original news article after showing a summary) Does the summary reflect the contents of the original news article well?	5.3	4.7
Q2	(show a summary after showing an original news article) Does the summary reflect the contents of the original news article well?	5.8	4.9

Table 8. Results of the Test 1 (Summary Receptiveness Test)

And, in the summary method test as Test 2, we attempted to evaluate whether a summary from the proposed method is better than other basic methods, namely top-positioned sentences and random sampling. We showed 20 randomly selected pairs of an original news article and a summary to each tester, excluding the pairs used in Test 1. Testers are not allowed to know about the name or method of these pairs. We also used the same 7-point measure for this test. As the final outcome, we achieved relatively higher score (Q1) than other methods (Q2 and Q3).

	Question	Korean	English
Q1	(show summary sentences from the proposed method after showing an original news article) Does the summary reflect the contents of original news article well?	5.3	4.8
Q2	(show summary sentences positioned at the beginning after showing an original news article) Does the summary reflect the contents of original news article well?	5.0	4.7
Q3	(show summary sentences from random sampling after showing an original news article) Does the summary reflect the contents of original news article well?	4.5	4.3

Table 9. Results of the Test 2 (Summary Method Test))

6 CONCLUSIONS

This paper has presented an user system interaction that provides keyword extraction, summary sentence extraction and a search engine for users to efficiently read news articles on the mobile devices like a smart phone, with limited computing resources. For effective summarization, we used the method of keyword extraction by *Binary Independence Model* and the query expansion technique by the pseudo relevance feedback.

Our summarization system showed the best performance among other statistical methods in both of single and multiple article summarization evaluations. In addition, user receptiveness tests verified that the proposed system can provide a better quality summary.

Acknowledgement

This work was supported by ATC (Advanced Technology Center) Program – "Development of Conversational Q & A Search Framework Based on Linked Data: Project No. 10048448 [tel: 10048448]" and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. NRF-2013R1A1A2009937).

REFERENCES

[1] Deray, K.—Simoff, S.: Designing Technology for Visualisation of Interactions on Mobile Devices. Journal of Computing Science and Engineering, Vol. 3, 2009, No. 4, pp. 218–237.

[2] Buyukkokten, O.—Garcia-Molina, H.—Paepcke, A.: Seeing the Whole in Parts: Text Summarization for Web Browsing on Handheld Devices. Proceedings of 10th International Conference on World Wide Web, 2001, pp. 652–662.

[3] Rahman, A. F. R.—Alam, H.—Hartono, R.—Ariyoshi, K.: Automatic Summarization of Web Content to Smaller Display Devices. Proceedings of 6th International Conference on Document Analysis and Recognition, 2001, pp. 1064–1068.

[4] Jones, M.—Marsen, G.—Mohd-Nasir, N.—Boone, K.—Buchanan, G.: Improving Web Interaction on Small Displays. Computer Networks: The International Journal of Computer and Telecommunications Networking, Vol. 31, 1999, pp. 1129–1137.

[5] Sweeney, S.—Crestani, F.: Effective Search Results Summary Size and Device Screen Size: Is There a Relationship? Information Processing and Management, Vol. 42, 2006, No. 4, pp. 1056–1074.

[6] Berger, A. L.—Mittal, V. O.: OCELOT: A System for Summarizing Web Pages. Proceedings of 23rd Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, 2000, pp. 144–151.

[7] Harabagiu, S. M.—Lăcătuşu, F.: Generating Single and Multi-Document Summaries with GISTEXTER. Proceedings of Document Understanding Conference, 2002, pp. 30–38.

[8] Matsuo, Y.—Ishizuka, M.: Keyword Extraction from a Single Document Using Word Co-Occurrence Statistical Information. International Journal on Artificial Intelligence Tools, Vol. 13, 2004, No. 1, pp. 157–169.

[9] Svore, K.—Vanderwende, L.—Burges, C.: Enhancing Single-Document Summarization by Combining RankNet and Third-Party Sources. Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 2007, pp. 448–457.

[10] Ko, Y.—Seo, J.: An Effective Sentence-Extraction Technique Using Contextual Information and Statistical Approaches for Text Summarization. Pattern Recognition Letters, Vol. 29, 2008, No. 9, pp. 1366–1371.

[11] Li, L.—Wang, D.—Shen, C.—Li, T.: Ontology Enriched Multi Document Summarization in Disaster Management. Proceedings of the 33rd International ACM-SIGIR Conference on Research and Development in Information Retrieval, 2010, pp. 819–820.

[12] Jo, T.: Representation of Texts into String Vectors for Text Categorization. Journal of Computing Science and Engineering, Vol. 4, 2010, No. 2, pp. 110–127.

[13] Intelligent Morphological Analyzer, 21th Century Sejong Project (in Korean). Availaible on: `http://www.sejong.or.kr`.

[14] Salton, G.: Automatic Text Processing: The Transformation. Analysis and Retrieval of Information by Computer. Addison-Wesley Publishing Company, 1988.

[15] Robertson, S. E.—Jones, K. S.: Relevance Weighting of Search Terms. Journal of the American Society for Information Science, Vol. 27, 1976, No. 3, pp. 129–146.

[16] Ko, Y.—An, H.—Seo, J.: An Effective Snippet Generating Method Using the Pseudo Relevance Feedback Technique. Proceedings of 30th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, 2007, pp. 711–712.

[17] Ko, Y.—An, H.—Seo, J.: Pseudo-Relevance Feedback and Statistical Query Expansion for Web Snippet Generation. Information Processing Letters, Elsevier Science, Vol. 109, 2009, No. 1, pp. 18–22.

[18] Wasson, M.: Using Leading Text for News Summaries: Evaluation Results and Implications for Commercial Summarization Applications. Proceedings of 17th International Conference on Computational Linguistics and 36th Annual Meeting of the ACL, Vol. 2, 1998, pp. 1364–1368.

[19] Kim, T.—Park, H.—Shin, J.: Research on Text Understanding and Retrieval/Summarization/Filtering (in Korean). Technical Report. Korea Research and Development Information Center, 1999.

[20] Supplementary material for the paper "Efficient Keyword Extraction and Text Summarization for Reading Articles on Smart Phone". Availaible on: `http://nlp.sogang.ac.kr/SummurizationTestDataSets.html`.

[21] The New York Times (U.S. Edition). Availaible on: `http://www.nytimes.com`.

[22] Steinberger, J.—Ježek, K.: Evaluation Measures for Text Summarization. Computing and Informatics, Vol. 28, 2009, No. 2, pp. 251–275.

[23] Salton, G.—Fox, E. A.—Wu, H.: Extended Boolean Information Retrieval. Communications of the ACM, Vol. 26, 1983, No. 11, pp. 1022–1036.

[24] Kim, J.—Kim, J.—Hwang, D.: Korean Text Summarization Using an Aggregate Similarity. Proceedings of 5th International Workshop on Information Retrieval with Asian Languages, 2000, pp. 111–118.

[25] Carbonell, J. G.—Goldstein, J.: The Use of MMR: Diversity-Based Reranking for Reordering Documents and Producing Summaries. Proceedings of the 21st Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, 1998, pp. 335–336.

Hyoungil Jeong received the B.Sc. degree in computer science at Sogang University in 2006, and M.Sc. degree in computer science and engineering at Sogang University in 2008. He is a Ph.D. candidate in the Department of Computer Science and Engineering at the Sogang University from 2008.

Youngjoong Ko received the B.Sc. degree in mathematics at Sogang University in 1996 and the M.Sc. and Ph.D. degrees in computer science at Sogang University in 2000 and 2003. He is Associate Professor in the Department of Computer Engineering at the Dong-A University from 2004.

Jungyun Seo received the B.Sc. degree in mathematics at Sogang University in 1981 and the M.Sc. and Ph.D. degrees in computer science at the University of Texas, Austin in 1985 and 1990. He was Professor at the Computer Science Department of Korea Advanced Institute of Science and Technology (KAIST) from 1991 till 1995. He is Full Professor in the Department of Computer Science at the Sogang University from 1995.