News to Go: Hierarchical Text Summarization for Mobile Devices

Jahna Otterbacher*
Department of Public and Business Administration
University of Cyprus
P.O. Box 20537
CY-1678 Nicosia, Cyprus
jahna@ucy.ac.cy

Dragomir Radev
School of Information and Department of EECS
University of Michigan
1085 South University Ave.
304 West Hall
Ann Arbor, MI 48109-1107
radev@umich.edu

Omer Kareem
Department of EECS
University Michigan
1301 Beal Ave.
Ann Arbor, MI 48109-2122
okareem@umich.edu

ABSTRACT

We present an evaluation of a novel hierarchical text summarization method that allows users to view summaries of Web documents from small, mobile devices. Unlike previous approaches, ours does not require the documents to be in HTML since it infers a hierarchical structure automatically. Currently, the method is used to summarize news articles sent to a Web mail account in plain text format. Subjects used a Web-enabled mobile phone emulator to access the account's inbox and view the summarized news articles. They then used the summaries to complete several information-seeking tasks, which involved answering factual questions about the stories. In comparing the hierarchical text summary setting to that in which subjects were given the full text articles, there was no significant difference in task accuracy or the time taken to complete the task. However, in the hierarchical summarization setting, the number of bytes transferred per user request is less than half that of the full text case. Finally, in comparing the new method to three other summarization methods, subjects achieved significantly better accuracy on the tasks when using hierarchical summaries.

Categories and Subject Descriptors

H.4.3 [**Information Systems Applications**]: Communications Applications

General Terms

Performance, Design, Experimentation, Human Factors

*This work was conducted while the first author was at the University of Michigan's School of Information.

Keywords

Mobile Computing, Summarization

1. INTRODUCTION

Wireless access to Web context using small devices (e.g. PDAs and mobile phones) continues to be in demand by a wide range of users. From checking one's email while on the road to keeping abreast of the latest news and financial information throughout the day, mobile Internet access is a promising addition to desktop Web use. However, this technology is challenged by the fact that Web pages are typically designed to be viewed using a stationary computer connected to the Internet through high capacity lines. To contrast, handheld devices necessarily have small screens and wireless bandwidth is limited.

In considering how to make Web browsing on small devices more efficient, previous research has taken two main directions. The first approach involves reformatting or adapting Web pages to be more appropriate for viewing on small screens, without altering the original content. For example, this might be done by splitting a given page into smaller parts (e.g. [3, 6]) or by delivering only the objects of a page deemed to be important (e.g. [15]) and eliminating non-essential items such as graphics. To contrast, another approach is to actually transform the content of Web pages to be more suitable for view on a small device, as suggested by Trevor and colleagues [12]. For instance, summarization of Web pages has been introduced as a means of presenting the user with only the most salient content expressed in the text on a page (e.g. [2, 14]), thus reducing the amount of information that needs to be transferred to and displayed on the user's small device.

Previously, we introduced a novel method for hierarchical text summarization, which is appropriate for use with mobile devices [11]. In addition, we implemented a system that emulates the use of a mobile phone for checking one's Web mail. In particular, the system allows the user to access hierarchical summaries of the items in his or her inbox. In the current paper, we present a user study that evaluates the hierarchical summarization method, as implemented in the Web-based system, against several baselines.

Our evaluation is task-based, and emulates the experience of a user who wishes to keep informed of current news events throughout the day using his or her Web-enabled mobile phone. More specifically, the subjects in our study used the hierarchical summaries and the baselines to answer factual questions about a set of news articles emailed to the user's inbox (e.g. by an online news alert service). We will show that the subjects achieved better task accuracy when using the hierarchical summaries, as compared to three other summarization methods. In addition, as compared to the case where full text articles are displayed to the user, our new method reduces the number of bytes transferred per user request by more than half. Even more promising is the finding that users take no longer to complete the tasks and are just as accurate as they are in the case were they are given full text articles.

The remainder of the paper is organized as follows. In Section 1.1, we will describe the hierarchical summarization method, as well as its current implementation in a Web-based system for checking one's email. After that, we will describe the setup of the study in Section 2. In Section 3, we discuss the variables studied in our experiments as well as the hypotheses that are of interest to us, while in Section 4 we present our analysis and comparison to previous work. Finally, we follow up with conclusions from our current study.

1.1 Hierarchical summarization

Our hierarchical summarization method is illustrated in Figure 1. Given an input document, hierarchical summarization operates in two stages. First, it computes the salience of each sentence in the document and ranks the set of sentences accordingly. In order to compute the salience of each sentence, we use a linear combination of four features: Centroid, which measures how similar a sentence is to the overall document [10], Position (of the sentence within the source document), Length, and SimWithFirst, which measures how similar the sentence is to the first sentence in the document (often the title or headline of the news article).

In the second stage, a tree is constructed from all of the sentences such that its root is the sentence with the highest salience and, given any sentence node with salience s at depth d, all sentences above that depth have a salience higher than s, while the salience of the rest of the sentences is below s. As shown in Figure 1, at the first level, the user is shown the set of sentences that have the highest salience score, s_1. The order of presentation is the same as that in the source document. Sentences having a salience score less than s_1 are initially hidden from the user. However, at each point where lower-ranking sentences have been hidden, the user can expand the summary and view the sentences at the next salience level, which in this example have a salience score of s_2.

This process is also illustrated in Figure 2, which shows the interface of the system used in our current study that we created using the DeckIt WAP mobile phone emulator. The left portion of the figure shows a view of the user's email inbox, to which a set of news articles has been sent. In the right side of the figure, the user has selected to view the summary for the fourth article in the inbox. Here it can be seen that sentences 3 and 8 from this article had a salience score of s_1 and sentences 4 through 7 have been hidden at the next level. The idea behind hierarchical summarization is that the user is first shown the most important sentences in an article, in order to get the gist of the story. If he or she finds the initial summary interesting or relevant, the user may "drill down" the details of the story by expanding the message. Thus, the motive is to save time, bandwidth and screen space by displaying the most important information first, while at the same time giving the user the opportunity to expand the finer details if desired.

2. STUDY SETUP

We conducted a user study in order to evaluate the effectiveness of hierarchical summarization, as implemented in our Web-based email system, in facilitating access to information in online news articles. As previously mentioned, we consider the scenario in which the user wants to keep current of news events throughout the day. We assume that the user subscribes to a service in which news articles are sent to her Web mail account, and she checks this account periodically from her mobile phone to keep informed of newsworthy events.

For the study, we collected five unique sets of 10 Associated Press (APW) news articles each. The 10 articles in each set were published on the same day and were sent to the Web email inbox. In other words, each set of articles represents a snapshot of current events for that particular day. The articles included both world news events as well as financial updates. In the experiments, subjects used our system to find answers to questions about a set of 10 news stories. Below we describe the tasks and treatments administered, as well as the experimental design and study execution in detail.

2.1 Tasks

For each of the five document sets used, we created an information-seeking task consisting of 10 questions (one question per each article in the document set). Following Morris and colleagues [7], we used multiple choice questions in which there were five possible answers (but only one correct answer) for each question. An example of a document set (i.e. the headlines of the 10 documents) is shown in Figure 3. The questions comprising each task concerned key facts about the stories or events described in the respective articles. For instance, the questions used in the task for the document set shown in Figure 3 are given in Figure 4. (The five possible answer choices for each of the 10 questions are not shown due to space limitations.) For all questions in all tasks, the answers were reported explicitly in their respective articles. Therefore, it was not necessary for subjects to use previous knowledge or reasoning in order to answer the questions.

2.2 Treatments

In completing a given task, a subject was assigned to one of six treatments (or system settings). In addition to the hierarchical summarization setting, we included settings at two extremes: the full text setting, in which subjects are shown the original news articles, as well as the setting in which nothing is given to the subjects other than the task questions themselves. This control setting accounts for the possibility that the questions themselves may contain some information about the news stories. Finally, we also administered three other summarization methods - a top 20% summary, a lead-based summary as well as a summary made up of randomly selected sentences. These methods are commonly used as baselines in text summarization evaluations (e.g. the Document Understanding Conferences [9]). The

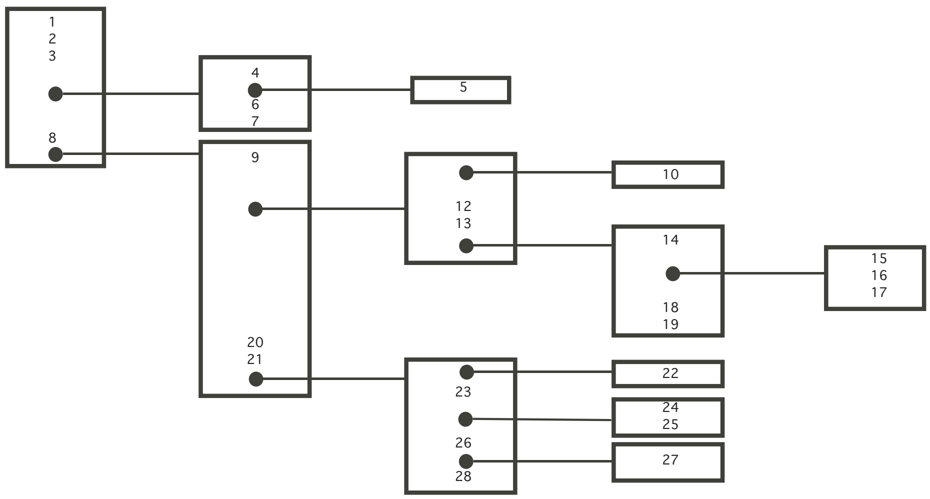

Figure 1: Hierarchical summarization.

Figure 2: The interface of WAP MEAD.

1. Vietnamese journalist awarded for devotion to free press
2. India's government faces budgetary woes
3. Malaysia Finance minister: Bad stats won't change growth forecast
4. TV, telephone, computer developments in Asia discussed
5. Australia clinches first wheat sale to Egypt mill project
6. Papua New Guinean leader to be first to greet Habibie
7. Protege of scandal-plagued president heads for runoff with rival
8. BC Britain Opening Gold
9. Dollar rises, stocks plunge in Tokyo trading prices
10. India denies it has plans another nuclear test With Pakistan-India

Figure 3: Example document set.

1. Who is Malaysian prime minister Mahathir's closet economic adviser?
2. How much wheat does Egypt import?
3. What is Kwinana?
4. Which stock index was down by 2.24 percent?
5. What was Doan Viet Hoat's occupation in 1976?
6. Who is Yashwant Sinha?
7. Where is Port Moresby?
8. Where is Irian Jaya?
9. What percentage of the vote in the Columbian presidential elections did the Liberal party win?
10. What is the name of a Japanese Car Producers' Organization?

Figure 4: Example set of 10 factual questions.

six treatments are described in Table 1.

Note that when using the system in each of the six treatment settings, for a given set of 10 documents, the user sees the same display first, which is the email inbox. The inbox shows a list of the 10 headlines of the articles in it. Therefore, Table 1 explains what is displayed on the user's mobile phone screen after he or she selects one of the news story headlines.

2.3 Experimental design and study execution

A total of 39 subjects was used in the study. They were recruited through an email sent to students studying information and computer sciences at our university. All subjects self-reported as native or near-native English speakers who were experienced Web users. Finally, they were paid for their participation in the study.

Although they were encouraged to complete all five of the tasks, the subjects were not required to do so (due to university research policies). Therefore, the researchers made sure that each of the five tasks and six treatment settings were assigned approximately equally often in the experiments. A balanced, incomplete block design was used, and the counts of each of the 30 possible document set-treatment pairings are shown in Table 2. In addition, the treatment and document set orderings were varied in order to prevent learning effects.

Before the experiments, the subjects were not given any information about the system that they would be using. They were informed that they would be participating in an information retrieval study and that its purpose was to examine how people search for information using a Web-enabled mobile phone. Finally, the subjects were told to answer the questions in each task as accurately as possible and were given unlimited time to complete the tasks.

Docset	T1	T2	T3	T4	T5	T6	Total
1	5	5	3	5	6	6	30
2	5	4	7	5	7	3	31
3	5	2	11	4	5	5	32
4	5	5	1	2	4	8	25
5	2	7	4	8	2	3	26
Total	22	23	26	24	24	25	144

Table 2: Counts of the document set and treatment pairings.

3. VARIABLES STUDIED AND RESEARCH QUESTIONS

In comparing the users' performance on the information-seeking tasks across the six different treatments, we examined the time taken to complete a task (recorded in minutes and seconds), as well as task accuracy (i.e. proportion of questions correctly answered, in which each question is either correct or incorrect). These are commonly used measures in extrinsic, or task-based evaluations of text summarizers [4]. In addition, we also obtained the number of requests made by the user, which also corresponds to the number of mouse clicks (or hits) in this case, as well as the total number of bytes transferred while completing a task. This information was obtained from the log file of each user's session (completing one task using one system setting). We then computed for each session, the number of bytes transferred per user request, in order to compare the efficiency of each of the methods tested. This measure gives us an idea of how much data has to be transferred to and displayed on the user's wireless device each time he or she interacts with the system.

The means of the three response variables across the six

Treatment	Description
Full Text	Shows the full text of each news article in the inbox
Hierarchical Summary	Nested summary showing the top 4 sentences, followed by the next 3 ranking sentences, for each article
Top 20% Summary	Displays the top 20% of sentences for each article
Lead-based Summary	Shows the first 4 sentences of each article
Random Summary	20% of the sentences in each article are chosen at random for inclusion in the summary
No Summary	No news articles or summaries given

Table 1: The six treatments used in the study.

Setting	Time (min.)	Task accuracy	Bytes per click
Full text	19.5	0.94	2674.5
Hierarchical summary	17.5	0.83	1206.0
Top 20% summary	16.1	0.63	1175.4
Lead-based summary	15.2	0.68	1295.4
Random summary	12.7	0.59	1208.2
No summary	2.9	0.32	0

Table 3: Mean time to task completion, task accuracy and bytes transferred per click under each setting.

settings are shown in Table 3. The subjects were most accurate on the information-finding tasks when using the setting in which they were shown the full text of the news articles, with an average task accuracy of 0.94. They also took more time to complete the tasks (an average of 19.5 minutes) than they did when using the article summaries. This finding is not very surprising, as in the full text case, the answers to the questions will always be available to the user, such that it is simply a matter of taking the time to find them. However, as will be shown in Section 4, the differences in time and accuracy are not significant between the full text setting and that in which subjects used hierarchical summaries to complete the tasks.

Another expected finding is that all of the summarization methods reduce the data transferred per user request, by more than half as compared to the full text setting. This is intuitive since, when using the summarization techniques, the goal is to prioritize information by ranking the sentences according to salience and to display the sentences incrementally in rank order. Finally, we can see that in the "no summary" treatment, where subjects answered questions about a document set without access to the documents or their summaries, the accuracy is very low (an average of 0.32). Therefore, there is no evidence that the questions themselves contain too much information about the news stories and we are not concerned about the task being trivial.

In the next section, we will concentrate on answering three research questions using the data from the user study:

1. **Are there significant differences between the five treatments (systems) when the effects of task difficulty are controlled?**

The means of the three response variables, time to task completion, task accuracy and bytes transferred per hit, which are shown in Table 3, appear to differ between the five systems. In addition, in assigning subjects to a given task (i.e. set of documents and questions to answer) and setting, we tried to ensure an approximately even distribution of task-setting pairing. Nonetheless, we want to investigate the effect of the five system treatments on the three response variables when the possible effects of the task are controlled. For example, it may be the case that some tasks were more difficult than others. Likewise, it could be possible that certain task and system combinations resulted in longer task completion times or lower rates of accuracy.

2. **Are there any significant differences in task performance and efficiency between the hierarchical summarization setting and the full text setting?**

If we establish, in investigating our first research question, that there is a significant system effect on the three response variables, then we should make pairwise comparisons between the five systems. In particular, as shown in Table 3, the highest mean task accuracy (0.94) occurs when subjects use the full text of the news documents to complete the tasks. To contrast, the accuracy when using the hierarchical text summaries of the news articles is slightly less, at 0.83. After that, we see a drop off, as the system with the next best accuracy, the lead-based summary setting, has a mean accuracy of only 0.68. Therefore, we will compare the hierarchical summary case versus the full text setting in order to see if the differences between them are statistically significant.

3. **Are there significant differences between the hierarchical summarization setting and the other three summarization methods?**

Finally, we wish to compare the three response variables between the hierarchical summarization setting and the three other summarization methods. As mentioned previously, these three methods are commonly viewed as baseline systems. As can be seen in Table 3, all four of the summarization methods reduce the number of bytes transferred per hit, as compared to the full text case. Therefore, we want

Response variable	Setting	Task	Setting*Task
Time	0.0033	0.0944	0.3714
Accuracy	0.0000	0.0549	0.0756
Bytes per click	0.0000	0.0023	0.4222

Table 4: P-values for predictors Setting, Task and their interaction for ANOVAs on each response variable.

	Difference	P-value
Time (min.)	2.0	1.000
Accuracy	0.11	0.645
Bytes per hit	1468.5	0.000

Table 5: P-values for the differences in the response variables between the full text and hierarchical summarization settings.

to investigate whether the new, hierarchical summarization method offers any significant advantages over the baseline methods in terms of the users' performance on the tasks.

4. ANALYSIS AND DISCUSSION

Below, we analyze the data collected from our user study in order to address the three research questions put forward in the previous section.

4.1 Setting effect when task is controlled

In order to examine if there is an overall setting (or system) effect, when controlling for the task administered to the subjects, we conducted an analysis of variance (ANOVA) for each of the three response variables, which were all approximately normally distributed. First, we removed the cases where subjects were given only the tasks with no source articles or summaries (a control setting), in order to consider only the differences between the five systems (where either the full text documents were given to the user or one of the four types of summaries). In each ANOVA, the predictors were the setting used, the task/document set used, and the interaction between the given setting and task. For the ANOVAs on each of the three response variables, Table 4 shows the p-values of the three predictor variables.

As can be seen in the table, the setting effect is highly significant in all three ANOVAs, even when controlling for the effect of task. In fact, at the 5% significance level, the effect of the task assigned was significant in only one case, when the number of bytes per click is the response variable. Likewise, at this level, the interaction effect between the task and the system assigned is not significant for any of the response variables. (However, at the 10% level, the interaction is significant in the ANOVA of the response variable "accuracy.")

Therefore, we can conclude that there are significant differences between the five system settings in terms of the average time to complete the information-seeking task, the average task accuracy, and the number of bytes transferred per mouse click, even when we control the effects of the tasks and the interaction between the setting and task administered.

4.2 Hierarchical summarization versus the full text setting

Having established that there are significant differences between the five system settings, we can now make pairwise comparisons between them, in order to see which systems are better than others, in the context of the current task. Post-ANOVA pairwise tests can be conducted using the Bonferroni method [8]. Table 5 displays the differences in the average task completion times, task accuracy and bytes transferred per hit, between the full text setting and that in which users were shown hierarchical summaries of the documents in the email inbox. In addition, the corresponding Bonferroni-corrected p-values are shown.

We can see that the differences between the two systems with respect to the average time taken to complete the task and the task accuracy are not statistically significant, having large p-values of 1 and 0.6, respectively. To contrast, the difference in the mean number of bytes transferred is highly significant, with a p-value of 0. The interpretation of these findings is that there is no evidence of significant performance differences on the task of finding answers to questions about a set of news stories between the two settings. However, the use of hierarchical summarization in delivering newsworthy information to a user's mobile phone reduces the number of bytes transferred to the wireless device each time the user interacts with the system.

4.3 Hierarchical summarization versus the baseline methods

The post-ANOVA pairwise comparison tests were also used to examine the differences in the response variables between the hierarchical summarization setting and each of the three baseline summarization methods. The statistically significant differences are shown in Table 6 along with their corresponding p-values. It should be noted that the difference in accuracy between the hierarchical and the lead-based summarization methods is not significant at the 5% level, but only at the more lenient significance level of 10%.

As can be seen, users achieved better task accuracy when using the hierarchical summaries, as compared to the other three summarization methods. On average, the users took 4.8 minutes less to complete the tasks when using the random summaries as compared to the hierarchical summaries. However, the low accuracy achieved using randomly-created summaries (average accuracy of 0.59 as compared to 0.83 in the hierarchical summary setting) confirms one's intuition that the randomly generated summaries are of a relatively poor quality. Therefore, we suspect that the shorter task completion times might reflect users "giving up" on a search task if they are unable to find the answers to questions after exerting significant efforts. In conclusion, while the hierarchical summarization method does not offer an advantage over the baselines in terms of the time taken to complete the information-seeking tasks, the users achieved significantly better task accuracy using the new method.

4.4 Relation to previous work

The previous work closest to ours is that of Buyukkokten and colleagues [1, 2], who introduced a technique called "accordion summarization." This is similar to our method in that the basic idea is to reduce the amount of data transferred to a mobile device by showing the user information incrementally. However, one major difference with our work is

Comparison	Response variable	Difference	P-value
Hierarchical vs. Top 20%	Accuracy	0.20	0.0010
Hierarchical vs. Lead-based	Accuracy	0.15	0.0660
Hierarchical vs. Random	Time	-4.8	0.0300
	Accuracy	0.24	0.0000

Table 6: Significant differences and their p-values in response variables between the hierarchical and baseline summarization settings.

that their method relies on HTML artifacts that denote page fragments such as paragraphs and lists, in first identifying Semantic Textual Units (STUs). Once STUs are identified, they are then subjected to micro-level text summarization. To contrast, our method does not require that documents are in HTML format, as a hierarchical structure is inferred automatically using perceived sentence salience.

In addition, Buyukkokten and colleagues conducted a small user study in which 15 subjects performed information-seeking tasks using a PDA emulator. However, the tasks administered to the subjects included both factual questions as well as locating particular pages on the Web, while our work focuses on finding the answers to factual questions in news articles. Their best summarization method, which first displayed keywords for a Web page followed by the most salient sentence, was shown to reduce the users' search time as compared to other summarization schemes. They do not report on the users' accuracy on the information-seeking tasks administered.

In another related project, Yang and Wang proposed a summarization method for small mobile devices based on the fractal theory [14]. In their approach, a skeleton summary is first generated for an input document, which is based on the document's structure. For example, an HTML news documents might be made up of sections, paragraphs, sentences, terms and then words. Once the skeleton of the document has been generated, finer details can be added, in creating a summary of the document. As in the work of Buyukkokten and colleagues, the proposed fractal summarization method relies on an input Web document being formatted in HTML in order to infer its structure. While Yang and Wang also applied their method to the task of summarizing news documents for delivery to mobile devices in [13], they did not conduct any task-based evaluations of their approach.

The tasks we assigned to users in our study were similar to those used by Morris and colleagues [7]. In their work, they examined the effects of extractive text condensing (or extractive summarization) on users' reading comprehension. In their experiments, subjects completed tasks taken from a GMAT exam, which consisted of sets of multiple choice questions. Treatments included being shown the full text of the corresponding passage, an abstract of the summary constructed by an expert or extractive summaries of varying length. They found no significant differences in reading comprehension (measured as task accuracy) between the full text case as compared to the settings in which users were given human-constructed abstracts or 30% or 20% extractive summaries. Similarly, in our work, we also found no difference in task accuracy between the full text and hierarchical summary settings in our experiments.

Finally, another recent task-based evaluation of summarization techniques was conducted by McKeown and colleagues in the context of the Newsblaster system [1] [5]. While their work did not concern summarization for mobile devices, their goal was to evaluate the usefulness of multi-document summarization in helping users "make better use of the news." In their study, users were given summaries of a set of news documents about a particular topic, and were asked to write a summary report describing the main facts of the story. Four treatment settings were investigated: the use of full documents with no summary, a lead-based summary, a summary produced by Newsblaster, and human-created summaries. Their results showed that using the system-produced summaries resulted in better reports than did the use of documents only or the lead-based summaries. In sum, the findings of previous studies clearly demonstrate that summarization is a useful tool that can reduce the amount of text users must read without hindering their comprehension of the key ideas expressed, and the findings of our current study also concur with this conclusion.

5. CONCLUSIONS

We presented a user evaluation of a novel method for hierarchical extractive text summarization that is appropriate for use with mobile devices. Our method was tested in the context of a Web mail system, which allows a user to access his or her inbox, to which a set of news articles has been sent. We tested the use of hierarchical summaries against the use of full text documents as well as three baseline summarization methods (top 20%, lead-based and random) on the task of finding answers to questions about the given news stories. We found that there was no significant difference in terms of task accuracy and completion times between the full text document and hierarchical summarization settings. In addition, the use of the hierarchical summaries reduces the number of bytes per user request by more than half.

To our knowledge, the current work is unique both in terms of the summarization method used (hierarchical extractive summarization), as well as the application and task proposed (that of using a wireless mobile phone to keep up-to-date on the latest news events and financial information). In addition, previous work in summarization for mobile devices has either not been evaluated extrinsically (e.g. [13]) or has been evaluated on a rather small sample of users [2].

Finally, the application used in the current study addresses a real information retrieval need. In particular, as more information is transmitted electronically and is easily available on the Web, the more people require tools to man-

[1] *http://newsblaster.cs.columbia.edu*

age this information in a timely and efficient manner. A clear example of a user who would rely on such a system is a professional who needs access to the most recently available, newsworthy information in order to make informed decisions, even when he or she is not in front of a desktop computer. The current work has shown that hierarchical summarization can enhance the experience of such users in using small, wireless devices by reducing the amount of data that needs to be transferred to the user's phone or PDA, without adversely affecting his or her comprehension of the information of interest. Therefore, we plan to deploy hierarchical summarization in a number of ways. In particular, we will link the system we implemented to our online news service, NewsInEssence[2]. While NewsInEssence already allows users to request to receive news updates and summaries via email, in the future, we may also offer them the option of receiving hierarchical summaries, so that they may read them on their wireless devices. In addition, in future work, we plan to extend our hierarchical summarization system so that it may be personalized by users, allowing them to summarize other types of texts that they need to be able to access while away from their desktop computers (e.g. email).

6. ACKNOWLEDGMENTS

This work was partially supported by the U.S. National Science Foundation under the following grant: 0329043 "Probabilistic and link-based Methods for Exploiting Very Large Textual Repositories" administered through the IDM program. All opinions, findings, conclusions, and recommendations in this paper are made by the authors and do not necessarily reflect the views of the National Science Foundation. The authors would like to thank the members of the CLAIR research group, Yang Ye, and the anonymous SIGIR reviewers for their feedback and comments on this work.

7. REFERENCES

[1] O. Buyukkokten, H. Garcia-Molina, and A. Paepcke. Seeking the Whole in Parts: Text Summarization for Web Browsing on Handheld Devices. In *Proceedings of the World Wide Web Conference (WWW10)*, Hong Kong, May 2001.

[2] O. Buyukkokten, O. Kaljuvee, H. Garcia-Molina, A. Paepcke, and T. Winograd. Efficient Web Browsing on Handheld Devices Using Page and Form Summarization. *ACM Transactions on Information Systems (TOIS)*, 20(1):82–115, January 2002.

[3] Y. Chen, W.-Y. Ma, and H.-J. Zhang. Detecting Web Page Structure for Adaptive Viewing on Small Form Factor Devices. In *Proceedings of the ACM Conference on the World Wide Web (WWW'03)*, pages 225–233, Budapest, Hungary, May 2003.

[4] T. Hand. A Proposal for Task-based Evaluation of Text Summarization Systems. In *Proceedings of ACL/EACL '97 Summarization Workshop*, Madrid, Spain, 1997.

[5] K. McKeown, R. J. Passonneau, D. K. Elson, A. Nenkova, and J. Hirschberg. Do Summaries Help? A Task-Based Evaluation of Multi-Document Summarization. In *28th Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, Salvador, Brazil, August 2005.

[6] N. Milic-Frayling and R. Sommerer. SmartView: Flexible Viewing of Web Page Contents. In *Proceedings of the 11th World Wide Web Conference (WWW '02)*, 2002.

[7] A. H. Morris, G. M. Kasper, and D. A. Adams. The Effects and Limitations of Automated Text Condensing on Reading Comprehension Performance. In I. Mani and M. T. Maybury, editors, *Advances in Automatic Text Summarization*, pages 305–323, July 1999.

[8] J. Neter, W. Wasserman, and M. H. Kutner. *Applied Linear Statistical Models, 3rd Edition*. Irwin, 1990.

[9] P. Over and J. Yen. Intrinsic Evaluation of Generic News Text Summarization Systems. In *Proceedings of the Human Language Technology Conference Workshop on Text Summarization (DUC 2003)*, Edmonton, Canada, May 2003.

[10] D. R. Radev, H. Jing, and M. Budzikowska. Centroid-Based Summarization of Multiple Documents: Sentence Extraction, Utility-Based Evaluation, and User Studies. Seattle, WA, April 2000.

[11] D. R. Radev, O. Kareem, and J. Otterbacher. Hierarchical Text Summarization for WAP-enabled Mobile Devices. In *28th Annual ACM SIGIR Conference on Research and Development in Information Retrieval (Demonstration session)*, Salvador, Brazil, August 2005.

[12] J. Trevor, D. M. Hilbert, B. N. Schilit, and T. K. Koh. From Desktop to Phonetop: A UI for Web Interaction on Very Small Devices. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology*, Orlando, Florida, 2001.

[13] C. C. Yang and F. L. Wang. Automatic Summarization of Financial News Delivery on Mobile Devices. In *Proceedings of the ACM Conference on the World Wide Web (WWW'03)*, pages 225–233, Budapest, Hungary, May 2003.

[14] C. C. Yang and F. L. Wang. Fractal Summarization for Mobile Devices to Access Large Documents on the Web. In *Proceedings of the ACM Conference on the World Wide Web (WWW'03)*, pages 225–233, Budapest, Hungary, May 2003.

[15] X. Yin and W. S. Lee. Using Link Analysis to Improve Layout on Mobile Devices. In *Proceedings of the ACM Conference on the World Wide Web (WWW'04)*, pages 338–344, New York, New York, May 2004.

[2] *http://www.newsinessence.com*