



Personal Information Access Using Proactive Search and Mobile Hypertext

Hyeju Jang, Seongchan Kim, Wookhyun Shin, and Sung-Hyon Myaeng,
Korea Advanced Institute of Science and Technology

Mobile phones and devices are evolving steadily, adding increasingly advanced capabilities. Today's smart phones, for example, have PC-like functionality with complete operating system software that provides a platform for developers. Even phones without an operating system often have

such functions as email, Internet navigation, music players, e-book readers, and GPS, in addition to the usual text messaging (short message service, SMS), camera, and phone. So now, besides holding telephone numbers and text messages, mobile phones also allow access to video clips, calendars, task lists, notes, and so on.

Given the increased volume and diversity of the information available on and through mobile phones, the capability to search for information on these devices will rapidly become significant. Used at any time and any place, mobile phones often serve as lightweight information-capture tools.¹ Mobile phones save information because users might want it again later. Studies have shown that 58 to 81 percent of Web pages accessed are revisits to pages previously seen, and that similar patterns of repeated access characterize Unix commands, library book borrowing, and human memory.²

Information access through a mobile phone presents both constraints and opportunities.

The display's small size limits the amount of information that the phone can display. In addition, mobile phones' limited text-input capabilities will likely affect how users search for information compared to their use of the traditional Web. Not only are advanced search features unavailable in mobile search, but the amount of effort required to enter a word on a mobile phone is more than double the effort required to enter a query on a standard keyboard.^{3,4} Because operating a mobile device is not usually the user's primary job,¹ the user operations must be simple and the number of operations minimized. On the other hand, such devices' mobility can provide contextual information—such as location and other personal information—that can contribute to a more sensible guide toward the right information.

To minimize the amount of information that the user must browse, the system must anticipate what the user will want at the next step in the given context and minimize the number of items to choose from. We call

A new framework combining proactive search with mobile hypertext lets mobile phone users avoid text entry and tailors search results to suit the context.

Related Work on Proactive Searching

Because mobile search has obvious limitations, several studies have addressed the concept of proactive search. Dik Lun Lee explains the proactive and context-aware nature of information push in mobile information retrieval and discusses user profiling methods for extracting content and location interests.¹

Karen Church and Barry Smyth also describe proactive search and introduce a search browser based on the concept.² The interface integrates user contexts in the form of temporal and location information cues, and it shows queries and results selections that other users made in the given location. Breaking away from the traditional query-box and results-list view of search, it presents a view of evolving search activities sensitive to user context and preferences.

Some researchers have attempted to provide personalized, context-sensitive search that is not only critical but also possible with a mobile device. In Gurminder Singh's work, for example, the mobile device selects urgent information on the basis of the user's contextual information.³ In most of these attempts, using contextual information means factoring the user's physical location into selecting and ranking search results.²

Robert Myers, Edgar Zapata, and Singh proposed a way of overcoming the difficulties of traversing information over different applications on a mobile phone.⁴ Their goal was to let users manipulate all data in a single space rather than going into multiple source applications. Similar to mobile hypertext (described in this article) this approach involves linking and grouping related items.

Stuff I've Seen (SIS) shares some goals with mobile hypertext.⁵ It attempts to make it easier to search

information seen in the past (which it considers personal information), regardless of application. However, SIS and mobile hypertext have several differences. First, SIS requires a user's direct input for a query, whereas our approach attempts to avoid that requirement with automatically constructed hypertext and proactive search based on the user's choice of an anchor text. Second, mobile hypertext deals with relatively shorter information items—SMS messages, memos, email, and photos, for example—whereas SIS indexes all the files on a desktop computer.

References

1. D.L. Lee, "To Find or To Be Found, That Is the Question in Mobile Information Retrieval," *Proc. Workshop Mobile Information Retrieval (MobIR 08)*, 2008, pp. 7–10; <http://www3.ntu.edu.sg/home/efstsai/pubs/MobIR08.pdf>.
2. K. Church and B. Smyth, "Who, What, Where & When: A New Approach to Mobile Search," *Proc. 13th Int'l Conf. Intelligent User Interfaces*, ACM Press, 2008, pp. 309–312.
3. G. Singh, "PIM for Mobility," *Proc. Personal Information Management SIGIR 2006 Workshop*; <http://pim.ischool.washington.edu/pim06/files/singh-paper.pdf>.
4. R. Myers, E. Zapata, and G. Singh, "Linking Information for Mobile Use," *Proc. 4th Int'l Conf. Mobile Technology, Applications, and Systems and 1st Int'l Symp. Computer Human Interaction in Mobile Technology (IS-CHI 07)*, ACM Press, 2007, pp. 607–613.
5. S. Dumais et al., "Stuff I've Seen: A System for Personal Information Retrieval and Reuse," *Proc. 26th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR 03)*, ACM Press, 2003, pp. 72–79.

this capability *proactive search*: predicting what the user might ask for before the explicit query and preparing a ranked list of information items for the given context. Proactive search uses contextual information to help overcome mobile devices' physical constraints. (See "Related Work on Proactive Searching" for a summary of other research on this topic.)

We can further facilitate information access on and through mobile devices by a hypertext mechanism we propose called *mobile hypertext*. With this mechanism, a mobile phone automatically interconnects various types of contents by identifying anchor terms and hyperlinks. Some of the hyperlinks, however, connect not directly to an information item such as a short message or email but to a dynamic node that contains a ranked

list of information item surrogates. The hypertext framework rarely requires users to enter text using a limited input method; instead, the user selects from a list of related information items. The dynamic aspect of hypertext is tied with the proactive search capability and facilitated by the manipulation of contextual information.

What Is Personal Information?

All the information stored in a mobile phone is personal, in that the device belongs to and is managed by a single person. Scheduling information and short messages received from a friend, for example, are both personal since they aren't usually made publicly available. However, because a mobile phone can serve as a central device to access information on the Web, we

must widen the scope of personal information that the phone's owner can search and manage beyond these obvious examples.

A recent study elaborates on the notion of personal information to include six types.⁵ Information is personal to "me" when it is controlled or owned by me; about me; directed toward me; sent by me; experienced by me; or relevant (useful) to me (see Table 1). The search method we propose assumes that the user will access all six types through a mobile phone, but we can extend the scope to include nonpersonal information that can be accessed with a search engine or Web browser when these become easily accessible with a mobile phone.

In our work, we assume personal information originates from the following sources: contact information,

short messages, email, schedules, notes, to-do lists, and multimedia objects (geographic maps, images, videos, and music). That is, these types of objects can be a hyperlink's destination. These sources consist of textual data, with the exception of multimedia objects, for which we assume user-provided tags are available. We leave Web search results for future research, although the framework and techniques we discuss can be used without difficulty when Web search can be customized for a small display.

What Is Mobile Hypertext?

Hypertext is text that contains references (hyperlinks) from an anchor text or an image area to another passage of text or image, which the user can access by clicking the hyperlink. The World Wide Web is the chief example of hypertext (see <http://en.wikipedia.org/wiki/Hypertext>). Wikis such as Wikipedia are another popular example; these collaborative sites include backlinks (incoming links to a Web page) and source tracking. Although the Web and wikis have made users familiar with using hypertext to search for information, in most cases the links are created manually. Content creators decide what information should be connected with other information.

Automatic hyperlink creation has been an active area for research, focusing primarily on anchor text identification and link detection.^{6,7} Some wiki systems appear to adopt a semi-automatic linking method with link disambiguation.⁶ Compared to the general problem of hyperlink generation, our work is somewhat different because the messages used for mobile hypertext are much shorter than Web pages or wiki documents. This led us to devise a method focusing on four aspects of term-to-term relatedness (described later).

Table 1. Six types of personal information.

Relation of information to "me"	Example
Controlled or owned by me	Messages in my email accounts, files on my hard drive
About me	Credit and medical information, histories of Web browsing, contact information
Directed toward me	Phone calls, short messages, other people's schedules shared with me
Sent (posted, provided) by me	Blogs, email, short messages, personal Web pages, published articles
(Already) experienced by me	Email I've read, Web pages I've browsed, other people's schedules I've seen
Relevant (useful) to me	Information (as in Web pages) I searched and found useful

In our work, mobile hypertext serves as a vehicle for search and internal management of personal information through hypertext that the system automatically constructs and maintains. The system makes connections between related information on a mobile phone by means of anchor text (such as "today's meeting" in a short message) and information objects (such as a meeting schedule). As does general hypertext, it provides a convenient way of navigating through related information regardless of information sources with a simple operation such as clicking a menu option, which is less vexatious than typing words in a mobile environment. Using hypertext to provide options for related information before it is asked for can also remind the user of important things.

To construct hypertext automatically, the system must recognize key entities to use as anchor text. We classify these entities into five categories: person or group names, place names, activity or event names, temporal expressions, and themes. Person or group names can link to contact information or to the name of the message's sender or receiver, for example. Temporal expressions (such as "tomorrow" or "July 17") are essential to understanding the context of an event or an activity (for example, "meeting" or "seminar"). Place names are of important use in identifying a meeting

place, a registered location, and so on. Themes include email subjects, functions, tasks, and so on (for example, "order" or "project A").

Destinations of hyperlinks from anchors include contact information, SMS messages, email messages, schedules, notes, to-do lists, and multimedia objects.

Mobile hypertext differs from general hypertext in that it consists of a static part, which is the same as general hypertext, and a dynamic part, which incorporates the notion of a proactive search taking into account contextual information such as the phone's location, time, and owner. In mobile hypertext, a word or a phrase (a term) in a text piece is connected to an information object, so that all the information objects on a phone can be represented as a directed graph, as in Figure 1.

The dynamic part of mobile hypertext includes a new type of temporary information object called a *ranked list*, which the system creates dynamically when it shows associated information to the user (see Figure 1). For example, when a new short message arrives at the user's mobile phone, the system identifies an anchor ("meeting," for example) so that it can search and rank related information objects in the background. It ranks information objects by both topical and contextual relatedness. The same anchor text in the same information

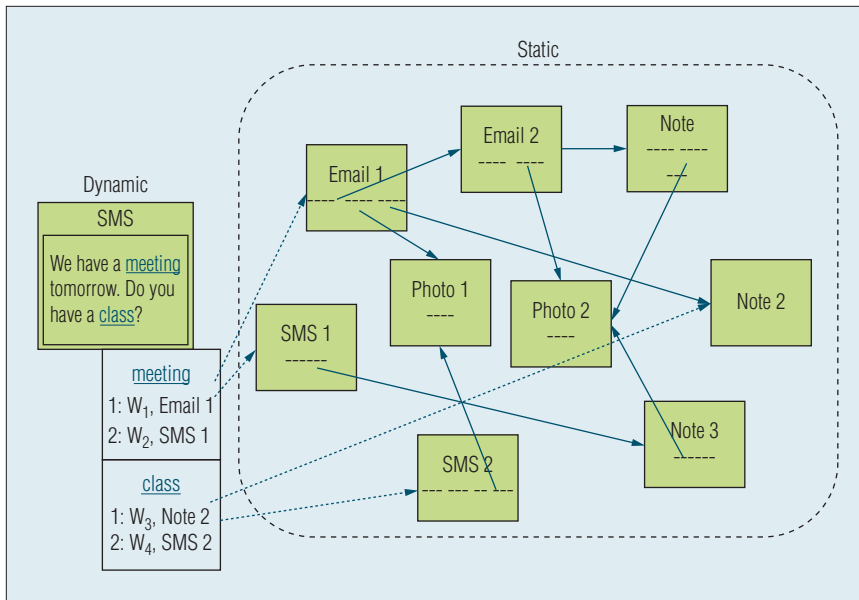


Figure 1. Mobile hypertext example. All the information objects on a phone are represented as a directed graph.

object sent to multiple users can result in different ranked lists depending on each phone's location, time, and owner. In essence, mobile hypertext contains ranked lists whose contents vary depending on the mobile phone user's context.

Automatic Construction of Mobile Hypertext

Automatically constructing mobile hypertext requires two major tasks: identifying anchor text candidates and determining information objects to link with an anchor. Thus, we need to apply some natural language processing and information retrieval techniques for text analysis and relatedness calculation. Figure 2 shows the flow of the processes required to generate a hyperlink.

Given a piece of text such as a short message, the system parses it with basic language-processing techniques such as tokenization, part-of-speech (POS) tagging, stop-word removal, and named-entity recognition (NER). The results are the entities classified into the five types mentioned earlier (person or group names, place names, activity or event names, temporal expressions,

and themes). In our system, this task is done by a rule-based classifier using a lexicon with the help of an NER module, although a machine-learning-based method could be used. The NER module detects person, organization, and location names. The classifier identifies activity or event names and temporal expressions on the basis of a lexicon, constructed by automatically extending a manually built lexicon with a thesaurus. Themes are the remaining nouns, which the POS tagger can identify.

Once the system has generated candidate anchors, the next step is to compare them with the terms in the existing information objects so that the computed relatedness can be used for ranking information items required for proactive search and hyperlink creation. The notion of relatedness can be interpreted in several different ways. In our work, we compare terms using four different aspects of relatedness: lexical, synonymous, paradigmatic, and temporal.

The final step before generating a hyperlink is context resolution and ranking, where the system considers time expressions and place information to determine whether a link

should be created and how to rank the related information objects. For example, even if the term "project meeting" in a new message matches several schedule items, the time expression "tomorrow"—which can be resolved to "June 19"—can select a particular schedule instance.

Hyperlink creation is bidirectional in that the new information object should be linked to and from some of the existing information objects. A set of candidate anchors from the existing information objects can be obtained easily with an inverted index. A hyperlink from a newly arrived information object to existing objects is called a *backward link*, whereas a link from an existing object to a new object is called a *forward link*. Because backward links lead to older related objects, they are useful for reminding the user of things that have come up in the past. On the other hand, forward links let the user understand how things have developed and propagated from some initial point.

When a new information object arrives at the mobile phone, the dynamic part of mobile hypertext comes into play in addition to the generation of new hyperlinks in the static part. The system initiates a proactive search, giving the user one or more anchor candidates as query options; from each anchor, the user can reach a ranked list of related information objects. At this point, the user's current situation or context becomes important. For example, when the original text that mentions a meeting doesn't contain a time expression, the proactive search ranks information about a meeting closest to the current time first in the related information list. Because mobile contents usually consist of a few necessary words, the user's current situation is an influential clue for ranking.

The system decides whether to create a hyperlink based on the ranking function that computes similarity between query and document terms. For this computation, we use an extension of *term frequency-inverse document frequency* (tf-idf), with semantic relatedness between terms calculated

based on lexical, synonymous, paradigmatic, and temporal relatedness. When the system encounters a semantically related term in a document, it increments both the tf and idf values by the term's semantic relatedness value (less than one). When a document contains more than one related term, however, the system uses the maximum relatedness value for idf.

After the system determines the similarity between a query and a document on the basis of term statistics, the final ranking reflects temporal relatedness and spatial relatedness in sequence so that the rank of a temporally or spatially related message goes up. The system computes time proximity in days between the text containing the anchor and the target text if both contain explicit time information. Spatial relatedness is based on lexical-level matching and is incorporated into the ranking function. In the current implementation, the inclusion of spatial relatedness is limited mainly because the problem of detecting a location or place name from GPS coordinates is beyond the scope of this work. A candidate target text containing the exact location mentioned in the source text is given the highest rank when all other relatedness criteria are equal.

Computing Relatedness between Terms

As mentioned earlier, our system computes relatedness along four

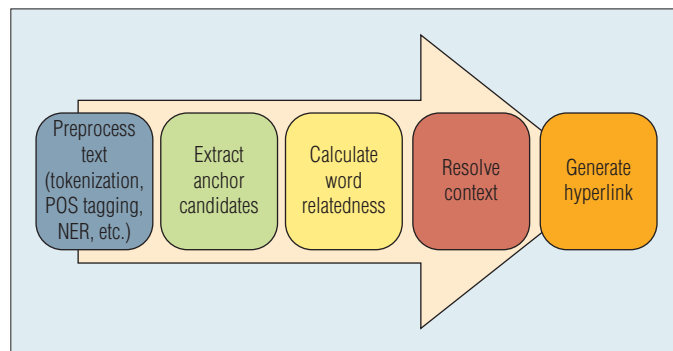


Figure 2. Automatic construction of mobile hypertext. The major tasks are identifying anchor text candidates and choosing information objects to link to each anchor.

dimensions. *Lexical relatedness* refers to how two terms are related on the surface based on word-to-word match. For example, “John Smith” is related to “Smith” and “John” at the pure lexical level, but, with a heuristic, more so to “Smith” than to “John.” This is because the family name “Smith” has a higher probability of referring to the person named “John Smith” than the first name “John” does. The system determines *synonymous* and *paradigmatic relatedness* values, on the other hand, using a thesaurus or lexical databases like WordNet. “Meeting” is synonymous with “seminar,” for example, whereas “bank” is paradigmatically related to “invest.” Paradigmatic relatedness can be found by traversing a subtree of a concept hierarchy or even using antonym relations. The system determines *temporal relatedness* for time expressions as in the example of “Meeting ... tomorrow”; this is essential for considering contextual information.

To compute relatedness, we basically make use of existing technology or a relatively straightforward method, although the problem awaits enhanced solutions, which we leave as future research. Following are descriptions of the methods in our current implementation.

Lexical Relatedness

Our system computes lexical relatedness chiefly using string matching

and dictionary lookup. It compares candidate phrases or words to compute the degree of overlap (for example, “food” versus “food market”). For named entities, the system uses an abbreviation dictionary (for example, to resolve “CA” and “California”) with heuristics like the relation-

ship between family and given names as mentioned earlier (“Smith” versus “John Smith”).

Synonymous Relatedness

To determine whether two terms have a synonymy relation, the current implementation consults WordNet. WordNet’s basic unit is a synset (a synonym set) that contains words and phrases with the same meaning. Synsets are connected with hypernym or hyponym relations and others such as antonyms. As long as two terms are found in the same synset, they are considered synonyms and have the maximum relatedness along this dimension. When a term is found in more than one synset, the system disambiguates it by using the other words in the same information object. Given mobile phones’ limited computational power and the fact that the vocabulary used on mobile phones is relatively small, however, we opted not to do sense disambiguation for the current implementation. Because our system is on an Android platform, we do not address the issue of keeping and accessing huge databases such as WordNet. Specialized lexicons and thesauri will have to be constructed just for personal management when the system is deployed to an actual phone.

Paradigmatic Relatedness

Two terms are paradigmatically related when they are found under a

reasonably narrow conceptual hierarchy or frequently found in proximity in a text corpus but aren't synonyms. For example, the words "agree," "delay," "argue," "majority," and "handout" are paradigmatically related because they are often mentioned together. Another example would be "subway," "passenger," and "crowded." Among several resources for finding paradigmatic relatedness—such as WordNet, *Roget's Thesaurus*, and Wikipedia for concept taxonomy, and the Web and various text corpora for co-occurrence statistics—we chose *Roget's Thesaurus* together with Jarmasz's method.⁸ Because the thesaurus doesn't contain newly coined terms that are often used in casual communications, we plan to also use the Web to compute paradigmatic relations.

Temporal Relatedness

Temporal relatedness is critical for personal information because most contents stored in a mobile phone are time-stamped. Because temporal expressions in the text part of information objects must be related among themselves as well as to the system time (for example, the message arrival time or the time a picture was taken), the system must compute their equivalence and temporal order. For instance, "tomorrow" and "Labor Day" must be converted to a day of the month and the year. Based on a set of rules and a lexicon, our system converts extracted time expressions into a canonical representation for a time point.⁹

Spatial Relatedness

Spatial relatedness is also important for context awareness in proactive search.¹⁰ To fully use the place names that can be identified from textual descriptions in information objects, however, the names must be related to the place where the information

object was created. Although GPS on a mobile phone can detect the current location, it's not an easy task to identify the corresponding place name. Consequently, spatial relatedness in our system is nothing but exact matching between place names at the lexical level.

Relatedness Value

Our system computes the final relatedness value by integrating the values from the four dimensions. Instead of taking a weighted sum of the four values, we take the maximum of the first three and see if it is over a threshold, because our goal is to ensure that the connection between a candidate anchor and an information object is strong. It would not make sense to create a link when there is weak evidence for each of the different types of relatedness. When "law firm" is the candidate anchor, for example, an information object including the text "I met the CEO of a company that produces golf clubs with a firm shaft" should not be linked even though "firm," "company," and "CEO" are likely to be related along the three dimensions.

Prototype System

We have developed a prototype system for mobile hypertext that includes the proactive search mechanism together with a user interface. Because existing mobile applications don't allow hyperlinks in their content, the first step was to build SMS, email, schedule, memo, and photo gallery applications on the Android platform using Android 1.5 software developer's kit. Next, we implemented the concept of mobile hypertext with logical links among information objects and a visual interface to show the hyperlinks.

Figure 3 shows an example session how mobile hypertext works in our prototype system. The first

interface (Figure 3a) shows the applications (services) running on the phone. In Figure 3b, the user notices that the list shows three messages; she selects the second message, to which she has not yet replied. In Figure 3c, the user reads the selected message and realizes that additional information is available for "class." Next to the anchor text are two icons that correspond to email and note. The system chose the term "class" as an anchor over "nlp" although it extracted both as anchor candidates. This is because our system selects an event or activity over a theme when they appear in the same vicinity and their scores are within a bound. This is to avoid too many links in a text. When the user clicks the anchor (Figure 3d), the display shows a list of titles for the related information objects, one email message and one note, below the message. The list was dynamically created by the proactive search mechanism. Contextual information, or time, affects the ranking here. Proactive search ranked the email message first because the time point indicated by "tomorrow" in the received SMS message is closest to the time indicated in the email text by "next week" coupled with the email's arrival time. When the user clicks the top-ranked item, the display shows the actual email text with several anchors and the type of information objects that can be reached through the associated hyperlinks (Figure 3e). These anchors and hyperlinks were created in the background when the message arrived. Although not shown, the system also created forward hyperlinks to the message. Clicking "homework," the user finds a new ranked list of information objects below the email message (Figure 3f).

Analysis of User Benefits

Because this work's novelty comes from the mobile hypertext framework

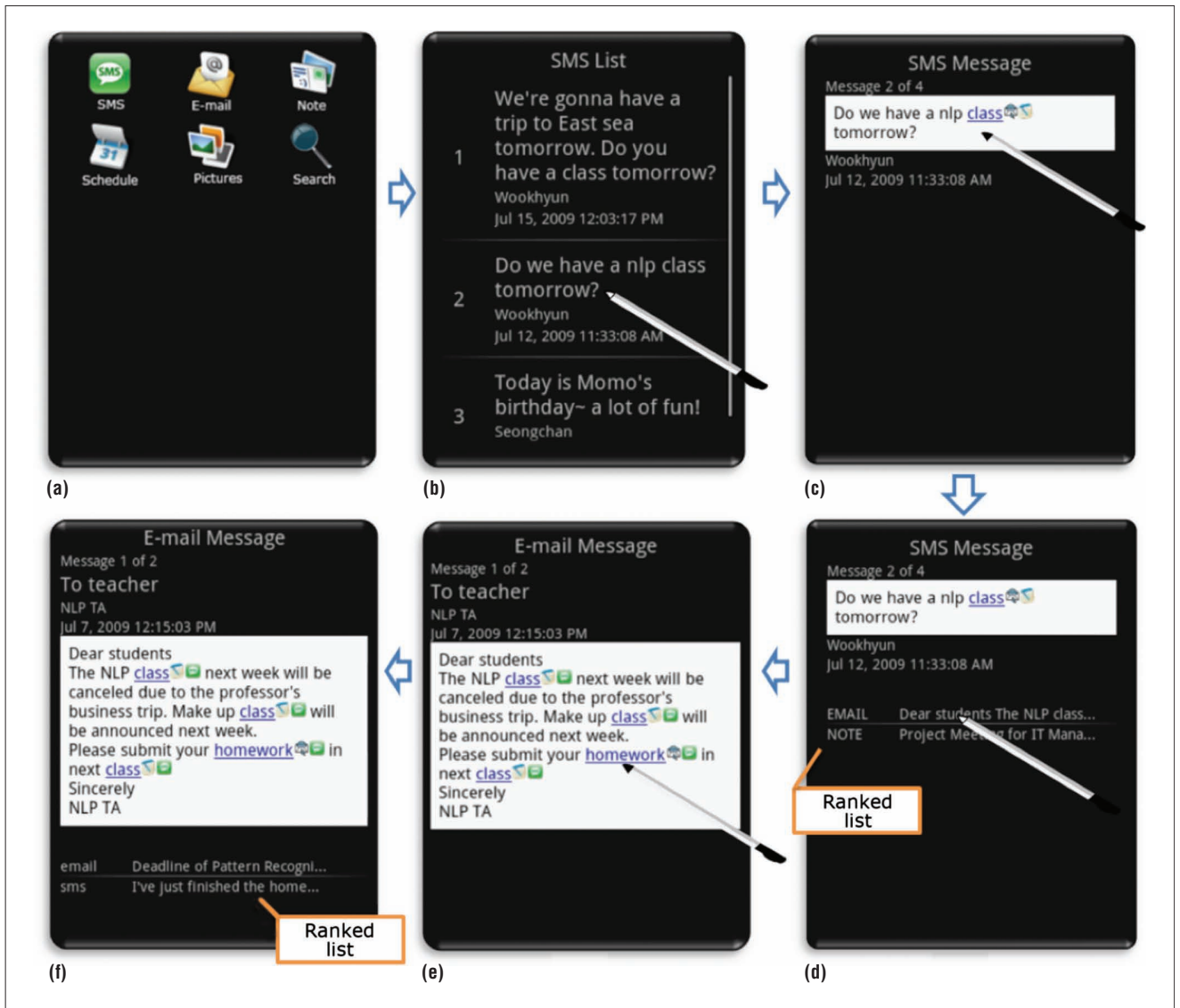


Figure 3. A scenario for mobile hypertext on the prototype system: (a) applications running on the phone; (b) the user selects a text message; (c) a hyperlink in the message indicates that additional information is available about “class”; (d) when the user selects “class,” the phone displays a ranked list of related information objects; (e) when the user selects the email from the ranked list, the related email message appears, containing additional hyperlinks; (f) selecting the “homework” hyperlink leads to a new ranked list of information objects.

and placing the proactive search concept on a mobile phone, we attempted to analyze its benefits in terms of reduced effort in information accesses and query entry, saved time, and reduced cognitive burden. With the purpose of finding evidence for the hypothesis that the interface using mobile hypertext helps users search personal information on a mobile phone, we conducted a usability test in October 2009 with a conventional

interface and the interface using the mobile hypertext framework.

Subjects

We recruited 20 participants, graduate and undergraduate students in IT-related departments in the university where we conducted our research. The participants’ average age was 24.4 years. Their average history using a mobile phone was 8.47 years; their average history using a PDA was

0.87 years. Participants assessed their skills using mobile phones as high (3.8 on a scale of 1.0 to 5.0). They were rewarded for their participation.

According to a questionnaire conducted after the experiment, the participants said that they seldom reuse the information on their phone once they check it for the first time. More specifically, although they reaccess SMS messages saved on a phone at least once a day, they “sometimes”

recheck other types of information such as memos and schedules. The chief reason they don't actively use their phones as information archives is that they have to go through many steps to find the information they want. Other reasons are inconvenience of operations such as text input, inability to use multiple windows, and lack of sufficient storage.

Baseline System

We built a baseline system whose interface features mimic those of popular phones the participants were already familiar with. To make the baseline system comparable to the mobile hypertext system, except for the hypertext aspect, we added some advanced features to the baseline system, making its interface look similar to that of a smart phone rather than a regular mobile phone. It included a shortcut to go to a different service directly at any time with a single click. In addition, four directional buttons allowed users to move vertically along the information items (for example, messages) within a service or horizontally to adjacent services directly (for example, from SMS to email with a right arrow) without having to return to the list of services. Moreover, we incorporated a keyword-search interface into the baseline system so users could enter keywords in a query window and obtain relevant results containing them. A search result would include a list of <service, information item> containing the exact query words. The baseline system's interface and functionality are identical to those of the proposed system except for mobile hypertext and proactive searching.

Experimental Settings

We built both the baseline system and the mobile hypertext system on an Android emulator platform. The

mobile hypertext system contains all the baseline system's features, including a keyword-search menu, plus hypertexts. The participants were allowed only to use the buttons and the keypad available on the emulator running on a desktop computer.

To reflect reality as closely as possible, the emulator contained 100 SMS messages, 100 emails, 8 notes, 40 pictures, and 100 schedule items for the experiment, which were edited from actual data on a student's mobile phone. The text was all in Korean to ensure the subjects could perform the tasks in a natural setting. Because the experiment's main goal was to test the usability of the concept of mobile hypertext, we adjusted the automatically generated links to exclude errors. This was a necessary step to ensure that the results would not be biased with errors.

Experimental Procedure

We first gave participants 20 minutes to read a short description of the data on a mobile phone and to look through the data as if the phone belonged to them. To further mimic the real situation in which people vaguely remember the existence of some information but not the exact location or details, we then gave them a 10-minute break. This would minimize the possibility of their memorizing the location and the exact information in the data set for the next step.

We then asked each participant to perform 10 tasks of searching for personal information on a mobile phone on the emulator. For each task, they had to find an information object containing information relevant to a specific question. We designed the 10 questions to have different levels of complexity and to force the users to use different services. An easy question to answer, for example, was "Where is the Noriplay performance

to take place?" Participants could obtain the answer from a single message stored in the email service through a keyword search or by going to the message directly if the participant remembered approximately when it arrived and who sent it. A more difficult task was to answer "What is the issue we'll have to talk about in the HCEP meeting this week?" Because the HCEP meeting was held weekly, and related information had piled up in the email and the schedule services, a simple search would have returned many items. Because this task required some temporal reasoning, participants found the answer more efficiently when they looked up the schedule first and linked to the email containing the announcement about the meeting.

To avoid the order effect, we divided the 20 participants into two groups of equal numbers and gave them two batches of tasks in an alternating fashion. For the first five tasks, group A used the baseline system while the group B used the mobile hypertext system. For the remaining five tasks, they switched the systems. Before the tasks, we gave participants an exercise task to familiarize them with the two interfaces.

For each information-seeking task, we measured the average number of steps, or click counts, and average time to reach a desired object. After participants had completed all the tasks, we gave them a survey to obtain their feedback. We asked them to rate the easiness and usefulness of the two systems, using a five-point scale with 1 being lowest (least useful/hardest) and 5 being highest (most useful/easiest).

Experimental Results

Figure 4 shows the average click counts and time required for users to obtain the desired object per task.

The charts demonstrate that the effort for finding information on the mobile hypertext interface is much less than on the conventional interface. The mean of the average click counts for all 10 tasks on the mobile hypertext interface is 2.31, much smaller than 7.84 on the conventional interface. For average time over the 10 tasks, the mean on the mobile hypertext interface is 19.06 seconds—again, much better than 56.80 seconds on the conventional interface.

Figure 5 compares user feedback on the two interfaces' easiness and usefulness. The users gave a very low easiness score to the conventional interface (that is, they found it hard to use). Although they did not feel that mobile hypertext was absolutely easy to use, they gave it a much higher score than the conventional interface. Both interfaces received higher usefulness scores, and the mobile hypertext interface's scores were surprisingly higher. We suspect that the participants found the mobile hypertext interface very useful for information seeking, partly due to its novelty, but that they did not always find it easy to use, partly because they weren't familiar with its operations. They indicated that the lack of desktop computer functionalities, such as multiple windows, was a major obstacle to its ease of use.

We observed a big difference in user behavior on the two interfaces that could be related to their judgments about easiness and usefulness. When users failed in searching using the conventional interface, they had to issue a new query or move to a different service (for example, from SMS to email). This is costly in terms of click counts and time, and it's a nuisance for users. However, with the mobile hypertext interface, when users arrived at irrelevant information, they simply pressed the "back" button and selected another possibly relevant

information item from the information list.

After the experiment, we let the participants describe the pros and cons of the newly introduced interface. Several participants noted that mobile hypertext saved time on information

searches. On the other hand, they pointed out that seeing too many links was bothersome and that the system should create a link only on the words a user would want to search with. Another interesting comment was that users could use phones as an information

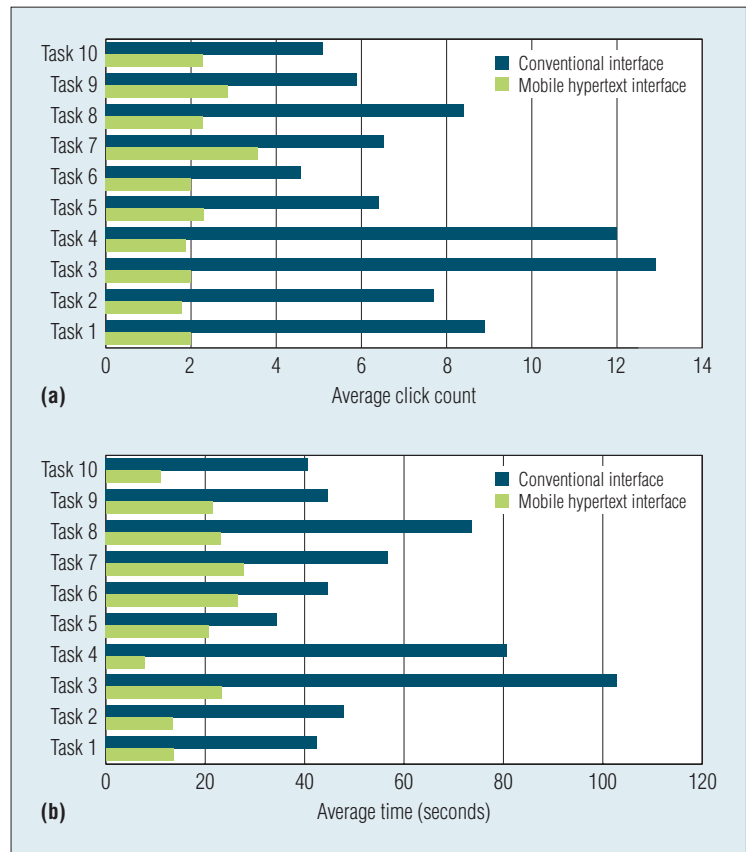


Figure 4. Mobile hypertext versus conventional mobile phone interface: average click counts and time per task.

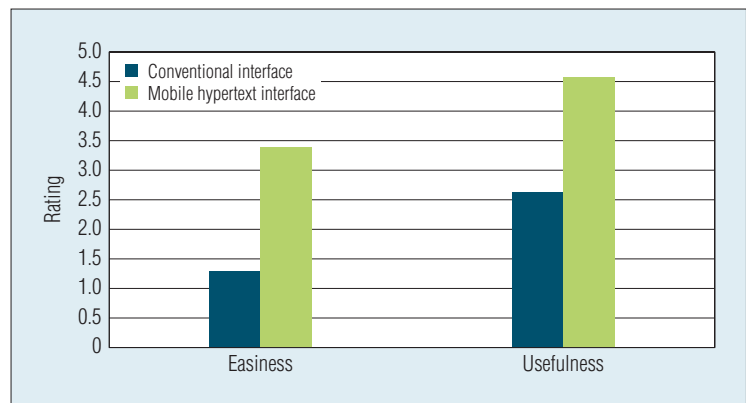


Figure 5. Mobile hypertext versus conventional mobile phone interface: user feedback on easiness and usefulness.

THE AUTHORS

Hyeju Jang is a researcher in the Information Retrieval and Natural Language Processing Lab in the Computer Science Department at Korea Advanced Institute of Science and Technology (KAIST). Her research interests are text mining and information retrieval in various domains. She has an MS from KAIST. Contact her at hyejujung@gmail.com.

Seongchan Kim is pursuing an MS in the Information Retrieval and Natural Language Processing Lab in the Computer Science Department at KAIST. His research interests include mobile information retrieval, mobile context-aware systems, and mobile personal information management. He has a BS from Chonbuk National University, Korea. Contact him at skim@kaist.ac.kr.

Wookhyun Shin is pursuing an MS in the Information Retrieval and Natural Language Processing Lab in the Computer Science Department at KAIST. His research interests cover opinion mining, Web 2.0 and society, and mobile information retrieval. He has a BS in computer science from KAIST. Contact him at wshin@kaist.ac.kr.

Sung-Hyon Myaeng is a professor in the Computer Science Department at KAIST. His research interests include information retrieval (especially mobile and contextual search), text mining, and context-aware computing, especially the interplay between human activities in the cyber and physical worlds. He has a PhD from Southern Methodist University. Contact him at myaeng@kaist.ac.kr.

archive more actively if they had a convenient tool to do so. We expect that mobile hypertext and proactive search will serve the purpose.

Although we've introduced the concept of mobile hypertext in conjunction with proactive search and shown its advantages through an experiment, further work remains, especially in enhancing the underlying techniques such as relatedness computation, hyperlink creation, and ranking methods for different types of information objects. In addition, when it becomes possible for mobile phones to infer the place in which an information object is created or sought, this contextual information will make mobile hypertext with proactive search even more effective. ■

Acknowledgments

Support for this research came from the Ministry of Knowledge Economy, Korea,


under the Information Technology Research Center support program supervised by the National IT Industry Promotion Agency; NIPA-2009-(C1090-0903-0008). Financial support for this work also came from a grant from the strategic technology development program 2008-F-047-02 of the Ministry of Knowledge Economy.

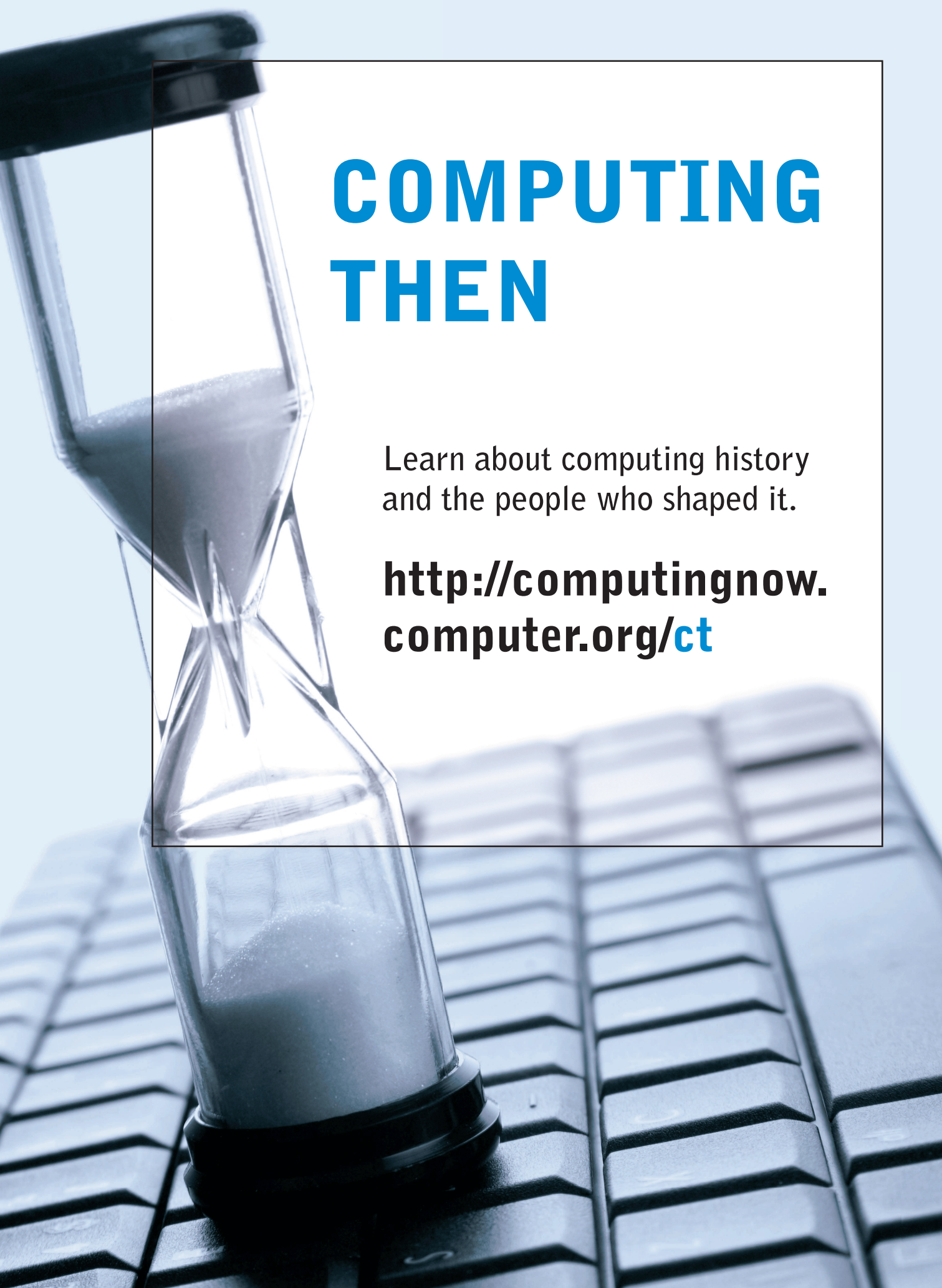
References

1. M. Tungare and M. Pérez Quiñones, "Thinking Outside the (Beige) Box: Personal Information Management Beyond the Desktop," *Proc. Personal Information Management Workshop, ACM Conf. Human Factors in Computing Systems (PIM 08)*, <http://pim2008.ethz.ch/papers/pim2008-tungare-et-al.pdf>.
2. S. Dumais et al., "Stuff I've Seen: A System for Personal Information Retrieval and Reuse," *Proc. 26th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR 03)*, ACM Press, 2003, pp. 72–79.
3. K. Church et al., "Mobile Information Access: A Study of Emerging Search Behavior on the Mobile Internet," *ACM*

Trans. Web, vol. 1, no. 1, 2007, article no. 4.

4. M. Kamvar and S. Baluja, "Deciphering Trends in Mobile Search," *Computer*, vol. 40, no. 8, 2007, pp. 25–62.
5. W. Jones, "How Is Information Personal?" *Proc. Personal Information Management Workshop, ACM Conf. Human Factors in Computing Systems (PIM 08)*, <http://pim2008.ethz.ch/papers/pim2008-jones.pdf>.
6. J. Gardner and L. Xiong, "Automatic Link Detection: A Sequence Labeling Approach," poster session, 18th ACM Conf. Information and Knowledge Management (CIKM 09), 2009, www.mathcs.emory.edu/~lxiong/research/pub/gardner09automatic.pdf.
7. S.R. El-Beltagy et al., "Linking in Context," *Proc. 12th ACM Conf. Hypertext and Hypermedia*, ACM Press, 2001, pp. 151–160.
8. M. Jarmasz and S. Szpakowicz, "Roget's Thesaurus and Semantic Similarity," *Recent Advances in Natural Language Processing III (RANLP 03)*, John Benjamins Publishing Co., 2003, pp. 111–120.
9. P. Kim and S.H. Myaeng, "Usefulness of Temporal Information Automatically Extracted from News Articles for Topic Tracking," *ACM Trans. Asian Language Information Processing (TALIP)*, vol. 3, no. 4, 2004, pp. 227–242.
10. K. Church and B. Smyth, "Who, What, Where & When: A New Approach to Mobile Search," *Proc. 13th Int'l Conf. Intelligent User Interfaces*, ACM Press, 2008, pp. 309–312.

 Selected CS articles and columns are also available for free at <http://ComputingNow.computer.org>.



COMPUTING THEN

Learn about computing history
and the people who shaped it.

[http://computingnow.
computer.org/ct](http://computingnow.computer.org/ct)