Smart Retrieval and Sharing of Information Resources based on Contexts of User-Information Relationships

Wai-Kwong Wing, Francis C.M. Lau, Cho-Li Wang
Department of Computer Science
The University of Hong Kong, Pokfulam Road, Hong Kong
{wkwing, fcmlau, clwang}@cs.hku.hk

Abstract

Information resources on their own present only the information they contain. The relationship between the resources and the users is usually neglected. By exploiting the relationship between the user and information resources in the form of context, which is established when the user accesses or acquires these resources, can help create smart mobile appliances. The metadata contained in a context is not merely data about data, but represents how the user and the information resource are related, which helps searching, provides clues to find related information resources, and facilitates information sharing with minimum manual effort. An electronic business name cards application is implemented to demonstrate the applicability of the idea.

1. Introduction

Searching for an electronic document, media file, or Web page, in a vast space of information is a difficult problem. The problem is now shifting from the non-mobile to the mobile world. Mobile information retrieval has attracted much attention from researchers [6].

There are many ways to perform searching, such as by keywords [2], historical search [17], and using metadata. In [17], a user's search history is categorized in order for the system to learn the user's past interests and to perform personalized Web searches. Many Web search engines are already performing keyword and metadata search to match user-specified keywords or synonyms with Web contents or metadata embedded in the contents. For example, Google's advanced search [1] lets users give a keyword for metadata matching, such as the title of a Web page. Searching by metadata, however, is not well researched when being applied in mobile environments.

Opportunities arising from the use of mobile devices call for smarter ways to find information. Because a mobile device accompanies and moves with the user [18], it is possible to take into account the situation (called "context") in which the user is in when running certain applications with the device. The problem we address in this paper is how to make a device smart by taking advantage of such contextual information.

2. Information Resources with Contextual Information

Traditionally, a piece of information is most commonly stored as a file with certain metadata [14] such as a file name. In many implementations, the metadata is a collection of name-value pairs [10]. The ways to find a saved information resource based on metadata, however, are limited. This is because the metadata is associated with the information, not the user. In fact, when an information resource is created or accessed, there exists a relationship between the user and the resource, such as the reason for creating or accessing the information resource. The relationship, if captured as part of the metadata (or user's context) associated with the information resource, may prove to be useful in many retrieval scenarios. To recognize and establish the relationship, we need a mechanism to attach contextual information to information resources when they are being accessed or created.

There are many definitions of "context" [4, 5, 7]. The context we discuss here is a relation between the information resource and the user. It differs from traditional metadata, which is data about data, in that it is the link that connects information and the user. Our context includes what, where, when, why, who and/or how the information resource is being assessed or saved. With such contexts, we could answer questions such as: Why did I get it? Where did I get it? Any saved documents that are related to my current task? Etc. Therefore, contextual information should be saved with information resources as users access and archive them. It could be used when dealing with such information resources in the future.

Context is a kind of content-independent metadata [3], since the information therein concerns more with the user than with the contents of the information resource. Lamming and Flynn [16], based on several psychological theories, concluded that people tend to be able to recall contexts that have physical associations. For example, it is easy to forget the contents of some target information resource, but not how the contents have come into existence.

Contextual information provides more dimensions for later retrieval in a mobile environment than in a non-mobile one. Because of the additional dimensions, searching is no longer restricted to keyword-based matching and scoring. Although an event of saving some information can happen with both mobile and non-mobile devices, contextual information in the mobile environment is much more diverse and richer, thus providing more possibilities in future retrieval. As shown in Section 3, the contextual information can be used to find related documents and for information sharing.

Traditional content-based metadata for a certain information resource is created by the author of the information resource or a delegate. The correspondence between the metadata, such as the title of a document, and the information is usually one-to-one. The metadata is relatively static without the need to change frequently. For user-based contextual information, however, the relation between the information resource and the user may change over time. There can be multiple pieces of user's contextual information attached to an information resource. Obviously, each user has his/her own contextual information. Such contextual information may come automatically from various sources such as the user's schedule, sensors, entities that interact with the user; it may also be created by the user.

3. Business Name Cards Application

We give examples on how contextual information would be useful in practical situations, via a an electronic business name cards application for PDAs. We have implemented a simple prototype, from which the following examples are derived. In addition to basic metadata about a name card [13], we create and store user contexts during exchanges of electronic business cards. Figure 1 shows a manager's PDA where a name card is being exchanged and a sample context created. When a name card is received, the name card (upper right) and the exhibitor's context (lower right) are retrieved. Table 1 explains the meaning of the XML tags in the figure. For our evaluation, we use imaginary name cards taken by an information technology manager who attends various technical and business conferences where exchanging electronic business cards (with exhibitors) is like a ritual. Without loss of generality, we used the vCard format for the prototype, as shown in Figure 1. The more important aspect is the saving of the context which can identify the name card (the "information_key" tag) and relate the card to the user. Here, the context of the exhibitor is received at the same time as the card is exchanged. The other contexts shown in Figure 2 come automatically from a location sensor/transmitter in the venue preset by the conference organizer (The "third_party_context" tag), as well as the manager's own schedule (The "owner_context" tag). The contexts are then merged and saved. The merged context provides many dimensions for retrieval and sharing. In the current implementation, when the name card is captured, the manager does not need to enter any data manually. This avoids any burden on the user. The context of the manager comes from some personal schedule whereas the context of the exhibitor and the conference organizer are preset manually. Allowing other auxiliary information to be entered dur-

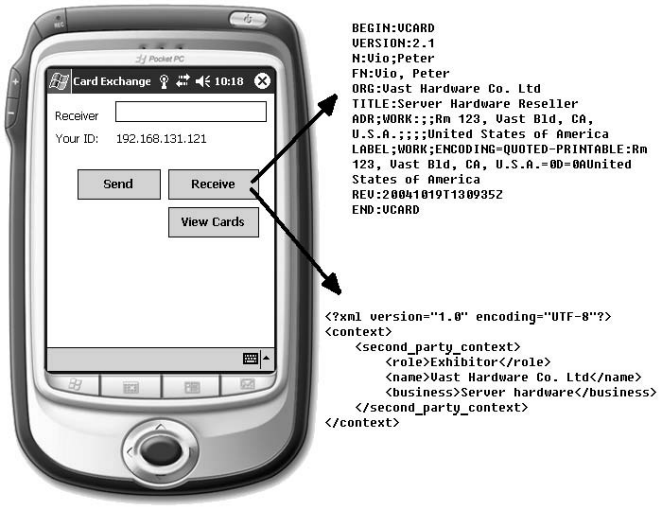

Figure 1. Exchanging a name card

```
<?xml version="1.0" encoding="UTF-8"?>
<context>
  <owner_context>
    <datetime>2004-09-30T09:23:02</datetime>
    <subject>Database</subject>
    <note>See the need for migration</note>
  </owner_context>
  <second_party_context>
    <role>Exhibitor</role>
    <name>Vast Hardware Co. Ltd</name>
    <business>Server hardware</business>
  </second_party_context>
  <third_party_context>
    <task>Conference</task>
    <type>Technical</type>
    <title>New Database Technologies, 2004</title>
    <organizer>Technics Data Co. Ltd</organizer>
  </third_party_context>
  <information_key>f89sdfy.vcf</information_key>
</context>
```

Figure 2. Merged context

Tag	Meaning
owner_context	Manager's context
subject	Subject line taken from manager's schedule
note	Body text taken from manager's schedule
second_party_context	Name card owner's context
role	Situation about how the name card is exchanged
name	Name of the exhibitor
business	Nature of business of the exhibitor
third_party_context	Conference organizer
task	Task of the conference organizer
type	Type of conference
title	Title of conference
organizer	Name of conference organizer
information_key	Key to find the corresponding name card (the information resource) in local storage

Table 1. Meanings of tags in Figure 2

Figure 3. Searching by context

ing or after capturing to enhance the smartness of the system is part of our future work.

3.1. Retrieval of Stored Information

There are many ways to search for a name card, such as keyword search by name and organization. Little research was done on searching by user context. Consider the manager again, where the name cards shown on his PDA are either sorted by name or by organization. Keyword searching is the only way to perform a direct search. Now with the presence of user contexts, the manager can sort the cards by event, as shown in Figure 3. Name cards can be grouped by different fields, such as by organization and by task. Within the group "task" name cards are grouped by contexts in which the cards were exchanged, such as "Conference". The figure shows that the manager attended a number of conferences including New Database Technologies, 2004. The name card of Vio Peter is finally reached after navigating through levels of contextual information. The various contexts related to the name card are displayed. Here, Vio appeared in the New Database Technologies as shown in the figure and the Security Conference which is not shown.

Keyword searching by past physical events is also possible. Entering "Conference" as a keyword allows the manager to search for all name cards taken from conferences. As the merged context is providing other dimensions for searching, common methods, such as search within results, can be adopted in context sensitive search. Such indirect search is based on the relationship between the user and the event encountered and between the event and the information resource. Without the event information, the linkage between the user and the information resource is unknown and searching without any information about the target is infeasible. In the manager's case, Vio is found through the searching sequence of {conference, database, hardware}. From this example, we see that the searching query matches the manager's intention, which is to contact someone whom he met in a conference and who sells hardware that runs databases. The person can still be reached even when the name of the person and the corporation the person works for are forgotten. Note that using content-based metadata without contextual information is unlikely to reach the same searching results. For example, Vio may sell hardware in general, so the keyword "Database" is not linked with Vio. By appearing in a database conference Vio provided the context which is to promote database related hardware. After that, searching by events is more appropriate when the user forgets all but the physical context.

3.2. Related Documents

Documents with matching keywords are often related. Similarly, documents with matching user context may also be related. In the example, all exhibitors in the database conference are related by "Database" in that particular context. Figure 4 shows that Thomas and Vio are related in that one sells database software and the other sells database hardware. When Vio is searched, Thomas would be shown as a related name card because they appear in the same conference.

Figure 4 demonstrated two functions, context searching and related name card suggestion. Here, entering "Conference" searches name cards that are taken from a conference. When reaching a name card, the related name cards

Figure 4. Interrelationships of resources

are also shown. Thomas and Vio were directly involved in the same conference, and therefore their relation is obvious. What is not so obvious is that Vio also participated in a security conference, as shown in Figure 4, which another exhibitor, John, has attended. Thus, Thomas is related to Vio in database and Vio is related to John in security. These relations together form a web of relationships. Searching then is not solely restricted to distinct information resources but extended to a series of related documents—in this example, from Thomas to John.

3.3. Information Sharing

Hyams and Sellen [12] found that there are many ways, such as organizing and structuring information in clusters, to enable information sharing. However, they tend to require manual work from the sharer (person sharing information) before the sharing, and that could be a burden to the users. If clusters are formed based on contextual information, manual work to organize the information can be minimized. Information seekers may search the shared information as usual, such as by keyword, metadata and context. In particular, contextual information could give hints on how information resources are used by the sharer.

Current search engines focus on information, rather than on how the information is being used. User's context represents some extra information [12] when sharing an information resource with other users. Look at the manager again. The manager has colleagues who also attended some conferences and exchanged name cards. Sharing name cards with only traditional metadata is difficult because the situations in the cards were exchanged are unknown. It is difficult for the seekers to determine whether certain name card is useful. For example, a name card exchanged with an exhibitor can be more useful than one from an attendee casually met. When both cards have similar content-based metadata, such as "hardware reseller", they are basically indistinguishable. In this case, the clusters are "Exhibitor" and "Attendee", where the name cards in the exhibitor cluster are perceived to be more useful and should be shared. In our implementation, the name cards and their corresponding contexts in the cluster are copied to a PC server for other managers to download. Name cards stored in the server are in vCard format with extension ".vcf" whereas contextual information is saved in XML format with extension ".xml". These files are grouped and copied in directories whose names denote the respective clusters. For example, referring to Figure 1, "f89sdfy.vcf", the name card of Vio, and the merged context, "f89sdfy.xml", are copied to the server's file system inside a directory called "Exhibitor", which is shared to other managers.

Besides, why a name card was taken could be answered by the stored context. As shown in Figure 2, "See the need for migration" is the answer. Seekers may retrieve the card based on why the sharer exchanged the card.

4. Other Applications

If the idea of attaching context to information resources is important, it should be applicable to a wide range of applications. We discuss some scenarios in this section.

There are a number of ways to annotate a digital photo, such as by capturing certain metadata at the time the photo is taken [20] or by suggesting categorized concepts [11]. Such annotation methods could be further enriched with user contexts. For example, a photo with metadata "Tom in National Park during Holiday Journey with Family" reveals more than "Tom in National Park". Although only Tom is in the photo, users can find the photo by entering "family journey" which is a stored context. Here, "Tom" and "National Park" are content dependent, whereas "Holiday Journey" and "Family" are the user's contextual information. The contextual information can be taken from the user's schedule recorded in the PDA, or from the context about who accompanies Tom.

URL bookmarks stored with various contexts help browsers to automatically categorize them. In the Web, most contents are information applications where their presentation and interactions are limited to the capability of a browser [21]. Bookmarking these contents, when uncategorized, form a long list of URL bookmarks with which users would find it difficult to locate one, or to determine a certain bookmark is in the list. Manual categorization helps, but requires manual work. Contextual information can enable automatic categorization. Consider an example, where a student has gathered various bookmarks in different situations, such as during lectures, in the library, or at home. Then, the student could find it eas-

ier to find a bookmark if the bookmarks can be displayed in different dimensions. The dimensions facilitates the division of the bookmarks into different clusters, such as by location (school, library, home), or by task (some lectures). The contextual information may be obtained from course enrollment system of the college.

A business manager downloads business news everyday to browse on a PDA. Each day there could be many pieces of news but only few would interest the manager. For the interested news, the manager will traverse down the news tree to read more. Such context about extended browsing of the news is saved, meaning the news may be important in a certain way. Later, when the manager searches from thousands of news articles stored locally or in a remote server and many pieces of news emerge as results, the searching results can be sorted by importance based on the manager's previous browsing behaviour. In addition, as the manager reads through the news and traverses various links inside, such accesses form a web of news articles relationships. Then in later searches, related news articles can be suggested to the manager. Here, the contextual information is the article browsing behaviour in terms of URL link traversal.

A research student may be in contact with thousands of journal articles. Currently, an article has metadata like paper title, journal title, contributors, keywords and categories. Since the student specializes in a particular research field, articles downloaded would belong more or less to the same keywords and categories. This makes sharing articles among schoolmates difficult. With user's contextual information, such as how and where an article is used, other students could use the information to determine the usefulness of an article. For example, an article referenced by a student in preparing a research project proposal would be more important than those that are not referenced. The context is how an article is used by the user.

5. Related Work

Forget-me-not [16] is a portable, context-sensitive diary targeting at information retrieval. It records data and context a user encountered when using ubiquitous devices. It is basically a memory aid [15]. The contexts in their implementation are limited to location, time, telephone conversation, e-mail, file or printing events and workstation computing tasks. Searching is then based on these contexts, such as the presence of someone, or something that was printed in some printers. Contexts are merely retrieval keys. In our system, we further exploit the context and use context to represent the relationship between information resources and the users who access and archive them. The resources form a web of documents relationships which can be used to suggest related resources, and as clustering criteria for sharing, which are not addressed in forget-me-not. For example, in our name cards application, name cards can be clustered for sharing, whereas in forget-me-not, name cards have to be arrived at separately.

Forget-me-not was implemented when the suitable hardware technologies to support their features were not quite available. For example, infra-red was used instead of wireless LAN, and central server was used to store data instead of their mobile devices. With better hardware now available, the potential of context is worth further investigation.

Lifestreams [8, 9] is a diary of electronic life, archiving a time-ordered stream of documents. The researchers point out that very few systems track when, where and why documents are created and deleted. Yet, they concentrate mostly on the time dimension when explaining their system. Documents are inserted into a stream and are sorted by their creation time. Users than pick out a document from a series of documents in a stream. Our research emphasizes on user's context, which includes but not restricted to the time dimension. Similar to forget-me-not, the use of the time-dimension context is for retrieval without considering related documents nor sharing.

Lifestreams provides incentives for automatically cataloging documents without the use of file and directory names. They did not aim at mobile devices. We believe such automatic cataloging feature is more useful in mobile contexts, and can help mobile users to reduce their mental load when accessing information in mobile devices, because of the reduced concentration when using mobile devices, the awkward input metaphor, as well as the richer and more diverse user context in the mobile environment.

Jimminy [19] stores personal notes with location, people, and subject. Old notes are suggested just-in-time based on similarity of current user context and the stored context in a wearable computer. The aim of the system is to evaluate the usefulness of current context to find old notes. Our system emphasizes more on the potentials of stored context. In their system, personal notes are related to past user context. How personal notes interrelate and how the notes may be shared are not explained. Furthermore, the context stored is relatively stable and would not change after the personal notes are reused. As we have suggested, as information resources are being used, new contexts arise and are attached to the information resources. For example, a name card could be used numerous times. It may be beneficial to attach the related context each time the name card is used, and the number of times a name card is accessed could be used to determine the importance of the person in question; such importance can be used to form a cluster for sharing.

6. Future Work

Capturing and attaching user contexts to information resources should not only be limited to business name card

systems. Therefore, there is a need to build a generic system architecture that suits these and other smart appliance systems. This involves how contexts could be captured, saved and made use of. How to handle and organize a vast among of accumulated user's contextual information and how to provide useful information based on the saved contextual information are the major issues to be resolved.

7. Conclusion

Traditionally, pervasive computing applications adapt their behavior based on current or historical user contexts. Little attention has been given to the relationships between various information resources and the users who accessed, archived, created or updated them. We propose to store the relation in terms of contexts which can then be used as a tool for several useful operations. Doing so, the system remembers the situation in which a user acquired or accessed the resource. The context can be used to help searching, to provide clues to find related information resources, and to facilitate information sharing. The idea of attaching user contexts to information resources could be applied to a number of applications. We implemented an electronic business name cards application to demonstrate the applicability.

Acknowledgements

The authors wish to thank the referees for their valuable comments, and Louis Chan for sharing his experiences in handling name cards. The mobile devices we used in the experiments are gifts from the HP (Mobile Technology for Teaching Grant Initiative 2004). This project is supported in part by an Hong Kong RGC grant ("Content adaptation for mobile computing").

References

[1] Google. http://www.google.com.

[2] Jakarta Lucene. http://jakarta.apache.org/lucene.

[3] V. K. A. Sheth and W. LeBlanc. Attribute-based access of heterogeneous digital data. In *Proceedings of Web Access to Legacy Data Workshop at the Fourth WWW Conference*, 1995.

[4] R. W. B. Schilit, N. Adams. Context-aware computing applications. In *Proceedings of Workshop on Mobile Computing Systems and Applications*, pages 85–90. IEEE Computer Society, 1994.

[5] G. Chen and D. Kotz. A survey of context-aware mobile computing research. Technical Report TR2000-381, Department of Computer Science, Dartmouth College, November 2000.

[6] G. G. Chowdhury. *Introduction to Modern Information Retrieval*. Facet publishing, 2nd edition, 2004.

[7] A. K. Dey and G. D. Abowd. Towards a better understanding of context and context awareness. In *Workshop on The What, Who, Where, When, and How of Context-Awareness, as part of the 2000 Conference on Human Factors in Computing Systems*, 2000.

[8] E. Freeman and S. Fertig. Lifestreams: Organizing your electronic life. AAAI Fall Symposium: AI Applications in Knowledge Navigation and Retrieval, November 1995.

[9] E. Freeman and D. Gelernter. Lifestreams: a storage model for personal data. *SIGMOD Rec.*, 25(1):80–86, 1996.

[10] D. K. Gifford, P. Jouvelot, M. A. Sheldon, and J. J W O'Toole. Semantic file systems. In *Proceedings of the thirteenth ACM symposium on Operating systems principles*, pages 16–25. ACM Press, 1991.

[11] K. Haase and D. Tamés. Babelvision: better image searching through shared annotations. *Interactions*, 11(2):18–26, 2004.

[12] J. Hyams and A. Sellen. How knowledge workers gather information from the web: Implications for peer-to-peer file sharing tools. In *Proceedings of Seventeenth Annual Human-Computer Interaction Conference*, 2003.

[13] R. Iannella. Representing vCard objects in RDF/XML. W3C Note, February 2001.

[14] G. J. F. Jones and P. J. Brown. Information access for context-aware appliances (poster session). In *SIGIR '00: Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 382–384. ACM Press, 2000.

[15] M. Lamming and D. Bohm. SPECx: Another approach to human context and activity sensing research, using tiny peer-to-peer wireless computers. In *Ubicomp*, pages 192–199, 2003.

[16] M. Lamming and M. Flynn. Forget-me-not: intimate computing in support of human memory. In *Proceedings of FRIEND21 Symposium on Next Generation Human Interfaces*, 1994.

[17] F. Liu, C. Yu, and W. Y. Meng. Personalized web search by mapping user queries to categories. In *Proceedings of the eleventh international conference on Information and knowledge management*, pages 558–565. ACM Press, 2002.

[18] A. Marcus and E. Chen. Designing the PDA of the future. *Interactions*, 9(1):34–44, 2002.

[19] B. Rhodes. Using physical context for just-in-time information retrieval. *IEEE Transactions on Computers*, 52(8):1011–1014, 2003.

[20] A. W. Risto Sarvas, Erick Herrarte and M. Davis. Metadata creation system for mobile images. In *MobiSYS '04: Proceedings of the 2nd international conference on Mobile systems, applications, and services*, pages 36–48. ACM Press, 2004.

[21] L. Roman, editor. *Designing Software for the Mobile Context: A Practitioner's Guide*, page 22. Springer, 2004.