# Efficient Mining of User Behaviors by Temporal Mobile Access Patterns



Juryon Paik
Pyeongtaek University

# Efficient Mining of User Behaviors by Temporal Mobile Access Patterns

Seung-Cheol Lee†, Juryon Paik†, Jeewoong Ok†, Insang Song† and Ung Mo Kim†

†Dept. of Computer Engineering, Sungkyunkwan University, S.Korea.

School of Information and Communication Engineering, SungKyunKwan University
300 Chunchun-dong Jangan-gu, Suwon, Kyonggi-do, 440-746, S. Korea

**Summary**

Ubiquitous computing offers various kinds of dynamic services to the mobile users with versatile devices at anytime and anywhere. In ubiquitous environment intelligent mobile agents are mandated to communicate with users and it is enabled by capturing interesting user's behavior patterns. Existing mining methods have proposed frequent mobile user's behavior patterns statistically based on requested services and location information. In this case some problems are caused because it was not considered that the mobile user's dynamic behavior patterns are usually associated with temporal access patterns. Therefore, ubiquitous computing provides the dynamic services with timely manner when user wants to get useful services information on time. In this paper, we propose a novel data mining method, namely *temporal mobile access patterns* that can efficiently discover mobile user's temporal behavior patterns associated with location and requested services. Furthermore, we present a novel data structure T-Map to store the temporal mobile access patterns. The advantage of our data structure compactly stores the user's behavior pattern according to location and service information in memory. Even though the information data sets require large shared memory when they are stored, our approach still provides fast access and consume less memory than other methods. The proposed technique in this paper works especially well for context-awareness data sets in mobile agent systems.

***Key words:***

*Data Mining, Temporal Mining, Access Pattern, Association rules, Mobile Agent System.*

## 1. Introduction

The rapid expansion of ubiquitous mobile computing [2] technology has created an unprecedented opportunity to gather and extract information from mobile agent systems. Further, how services and access methods can be provided to users in ubiquitous mobile computing environment is critical problems recently. Therefore, effective modeling the behavior patterns of users in mobile agent systems is becoming very important. Effective modeling the behavior patterns of users in mobile agent systems benefits not only the users in smart access by caching or prefetching [1][5] but also the mobile service providers in financial profit like advertising [6].

In the ubiquitous mobile computing environments, the mobile users may request diverse kinds of services and applications by mobile devices from arbitrary locations at any time via on networks. Obviously, the behavior pattern, in which the location and the service are inherently coexistent with temporal associated, of mobile users becomes more complex than that of the traditional mobile agent systems. To assist the user get interested information on time is one of the promising applications, especially in mobile agent systems.

This paper is aimed at modeling how to build mobile user's behavior pattern with a temporal association rule [3] [11] In mobile agent systems. The mobile agent system consists of semi-structure data like XML data. Temporal association rules can be used to decide the next likely user's request services based on significant dynamic correlations. In the past, sequential association rule [9] [12] have been used to capture the co-occurrence of user's sequential movement pattern in mobile web systems domains. Episodes were designed to capture significant patterns form sequence of events. However, these models were not designed for the user's temporal movement patterns in mobile agent systems. And the deficiency of existed studies is that they considered only one of the characteristics, i.e. location associated with requested services. Obviously, both movement and location or service requested with the temporal association rule should be considered simultaneously in order to discover complete information of user behavior patterns when the user request services. As a result, it remains an open question how to discover the mobile user's interesting movement behavior patterns by temporal mobile access patterns based on temporal association rule for user and services provider.

This work was supported by grant No. R01-2004-000-10755-0 from the Basic Research Program of the Korea Science & Engineering Foundation.

In the ubiquitous mobile computing environment, a novel data structures and methods have been needed in order to achieve communication, transmission and interchange of information and for mobile user, the mobile user wish to get interesting services at real time with the current location by the mobile devices. To this end, we propose a novel data structure, namely Temporal Mobile Access Patterns (T-Map), and a novel mining method based on a given the novel data structure which is temporal mobile access patterns for discovering the behavior patterns such that suitable services can be predicted and recommended to users. The novel data mining method is proposed for efficiently discovering the user's behavior patterns with temporal association rule which are composed of the user's movement access pattern with requested services and temporal data sets. In this work, we consider the following scenario as a motivating. Assume that we analyze log files at some special server that considerable number of people in Seoul exhibit following behaviors: If some person arrives at university area (suppose there are a lot of university) then we recommend her useful information, such as currently opened library information at the university based on most people being used the service information at current time. She would move to nearby opened library for studying or research books. It is mandatory that the recommended of the service information for opened library has to be based on strong confidence.

The main contribution of our approach is to extract efficiently user's behavior patterns at real time. In additional, the contributions of the novel data mining method, T-Map, is useful for decision support system and match-making of web service as well as service discovery in intelligent ubiquitous middleware for user's request information and context-awareness information.

The remainder of this paper is organized as follows. In Section 2, we briefly review some related research. Section 3 describes the system architecture in details. Section 4 explores a novel algorithm Temporal Mobile Access Pattern, namely T-Map and a novel data structure in order to the mobile access patterns with time-related. Section 5 concludes this paper and discusses the future direction of our research.

## 2. Related work

Efficient algorithms for finding frequent itemsets or sequences in very large itemset or sequence databases have been one of the key success stories of data mining research. One of the early computationally efficient algorithms was Apriori (Agrawal and Srikan, 1994), which finds frequent itemsets of length l based on previously generated (*l*-1)-length frequent itemsets. The Mining sequential patterns (Agrawal and Srikan, 1995; Agrawal et al., 1996) proposed to find frequent pattern. The Frequent Patterns without candidate generation algorithm (Han and Pei et al. 2000) extended the Apriori method to find frequent patterns with a novel data structure, called Frequent Pattern tree (FP-tree). The FP-growth algorithm (Han et al., 2000) that combines projection with the use of the FP-tree data structure to compactly store in memory the itemsets of the original database. the basic ideas in this algorithm were recently used to develop a similar algorithm for finding sequential patterns (Pei et al., 2001). An entirely different approach for finding frequent itemsets and sequences are the equivalence class-based algorithms Eclat (Zaki, 2006) and SPADE (Zaki, 2001) that break the large search space of frequent patterns into small and independent chunks and use a vertical database format that allows them to determine the frequency by computing set intersect-tions. However, these methods did not consider temporal data set so that they cannot be applied directly for providing efficient service to user in ubiquitous mobile computing environment.

In recent years, the temporal data mining method has been studied for extracting useful knowledge from temporal data sets. The research of temporal data mining has been proposed the Generalized Sequential Pattern, named GSP (Srikan and Agrawal, 1996) and the SPIRIT method. The GSP algorithm extended the Apriori-like level-wise mining method with time constraint to find frequent patterns in sequential databases. And the SPIRIT method proposed frequent sequence patterns by regular expression constraint. More recently, the research of related temporal association rule has been studied for cyclic association (B. Ozden, S. Ramaswamy, and A. Silberschatz. 1998) and temporal pattern with association rule by calendric association methods (X. chen, I. petrounias and H. Heathfield. 1998). The problem of finding frequent patterns in mobile agent systems has been extended to that of finding frequent user behavior patterns in mobile web systems (Vincet S. Tseng, and Kawuu W. Lin. 2005). However they did not consider temporal association rule in mobile agent systems. As a result, it remains an open question how to construct the best association rule based on sequential and temporal association rules for mobile user.

## 3. System Architecture

In this Section, we propose the system architecture for the proposed data mining mechanisms. Before going into the details, the intelligent middleware system architecture associated with mining mechanisms and the data mining mechanisms are recapped briefly.

### 3.1 Intelligent middleware platform with mining module.

In [7] J. Paik et al. improved intelligent middleware platform, CALM with the actual usage of mining algorithm, esp. XML mining. To the purpose, they developed the auxiliary albeit import module, context data mining module, whose goal is to find important but implicit context information for accurate decision-making.

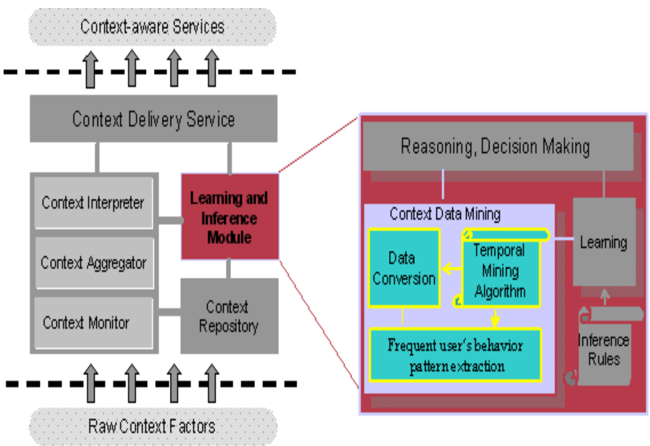


Fig. 1. Bird's eyes view of middleware platform and its mining module.

In this paper, we adopt their module for extraction user's behavior patterns and modify the algorithm to … our method. Fig. 1 presents how the original module for decision making has been altered. The left side of the figure depicts a simplified ubiquitous middleware structure in functional blocks for context-awareness. From the figure we can see that context factors are monitored, aggregated, interpreted, and finally both stored into storages and provided for the learning and inference module. The learning and inference module plays an absolutely key role in the middleware by deducing new and relevant information to application and user from various sources of context data. The aim of the module is to support the process of decision making for the adaptation at the middleware. However, the accuracy of learning and inference module can be degraded by the flood of useless and most of all, hidden context information. Therefore, the context data mining module is added an auxiliary module. This module is used to extract the mobile user's access pattern with time-related. Moreover, the auxiliary module is provided request information to user at real time associated with middleware.

### 3.2 System architecture for temporal mobile access pattern

In this subsection, we represent the system architecture associated with the context data mining module in details for mining temporal mobile access pattern. It is conducted to extract mobile user's useful access patterns with time-related. Fig. 2 describes the data mining system architecture for temporal mobile access pattern mining mechanisms. The work-flow of the system is divided two phases. This systems form that the logs for users' movement and users' service requests associated with temporal association can be stored in different databases. The first phase of system architecture, *Data integration phase*, is to collect and integrate users' logs into one dataset for efficient access and the temporary user information like current location to manage requests from subscribers. For this phase, the attributes related to user's temporary service requests will be extracted from the dispersed log files and joined to from an integrated log file by using the user's service identifier as the key. *Mining phase* is deployed novel data mining method to discover the frequently temporal mobile access patterns (T-Map) from the integrated log dataset. Consequently, the best recommended results are returned associated with temporally user's access pattern to the mobile agent. The mobile agent system is supported by both movement and service requests pattern of users with proper time information, which makes the system powerful and accurate.

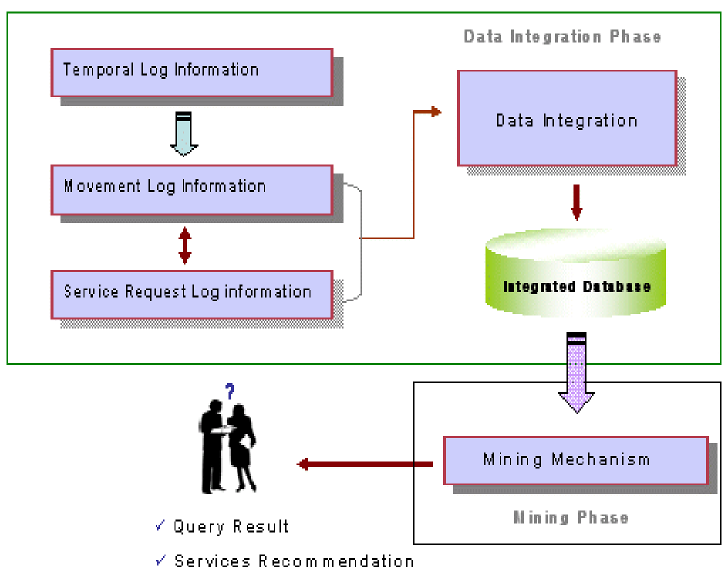


Fig. 2. System architecture for temporal mobile access pattern mining mechanisms

## 4. Mining of Temporal Mobile Access Patterns

In this section, we first give the formal definition to extract user's mobile access pattern with temporal association rule and propose a novel algorithm named T-Map that can discover the user's temporal movement behavior patterns related to location and service requests. Then, we explore a novel data structure. Describing that T-Map is combined user's temporal mobile access patterns to compactly store in memory.

### 4.1 Problem statement

A temporal database typically stores relational data with include time-related attributes. These attributes include (a) timestamp(s) of access patterns which one of the attributes has different semantics [10]. In this subsection, we consider two sets L and S with timestamps T, namely location set, service set, respectively. With each element *l* in L and s in S, we form an ordered pair $f$=($l$, $s$). Let $F$ be the set of all ordered pairs and it is denoted as

$$F = L \times S = \{all\ pairs\ of\ (l:s) \mid l \in L\ and\ s \in S\}$$

Let Mp = <($T_1$ ($f_1$, $tp_1$) ($f_2$, $tp_2$) ($f_3$, $tp_3$) … ($f_m$, $tp_m$)), ($T_2$ ($f_1$, $tp_1$) ($f_2$, $tp_2$) (f3, $tp_3$) … ($f_m$, $tp_m$)) … ($T_m$ ($f_1$, $tp_1$) ($f_2$, $tp_2$) ($f_3$, $tp_3$) … ($f_m$, $tp_m$))>, where $T_i$ is a timestamp during certain time interval, $f_i$ is the ordered pair, and $tp_k$ indicates the order of each $f_j$ (tp is implicit information). Mp describes mobile access pattern with time-series [8] (sequences of ordered access pattern with time-related). The predefined timestamps for the paper is illustrated in Table 1.

Table 1. An example of predefined timestamps $T_i$

| Timestamps | Time Interval |
|---|---|
| T1 | 06:01 ~ 10:00 |
| T2 | 10:01 ~ 14:00 |
| T3 | 14:01 ~ 18:00 |
| T4 | 18:01 ~ 22:00 |
| T5 | 22:01 ~ 02:00 |

For the simplicity, we set the time interval every four hours. The timestamps T1 is gathered user's occurrence access patterns from 06:01 to 10:00. T2, T3, T4 and T5 are also collected user's occurrence access patterns from 10:01 to 14:00, 14:01 to 18:00, 18:01 to 22:00, and from 22:01 to 02:00, respectively. Based on the set F and Table 1, the table of mobile access pattern, Mp, is generated. We construct the Mp, mobile access pattern, as shown in Table 2. To insert the first access patterns of user ID 100 into the table, the following process is constructing mobile access pattern (MP): i) Mobile user's current location with time-related into the integrated log to record the location and temporal information. Then, the user's location with time-related information is compared another user's location log data to find the same as current location information which occurred the same time in history movement log data set. The same as currently location information is extracted in history movement log data set. ii) The service request log is stored mobile user's current request service information. And then, the request service is also compared anther user's requested service information to find the same as requested service which occurred the same time in history service request log data set. Then, the same as request service information is extracted in history movement log data set. iii) Finally, The extracted location and request service information is integrated as shown Table 2, and the integrated information, location and request service, is stored in the integrated database. Therefore, we can discover the location based service with time-related at the same time. Table 2 shows the completed Mp table (mobile access pattern table).

Table 2. An integrated log data sets

| User ID | Timestamps | Access Pattern |
|---|---|---|
| 100 | T1 | < (O:1) (B:2) (P:5) (L:8) > |
| | T2 | < (A:1) (D:1) (E:2) > |
| | T3 | < (O:2) (B:2) (E:1) > |
| | T4 | < (C:4) (F:2) > |
| | T5 | < (C:2) (B:2) > |
| 101 | T1 | < (O:1) (B:3) (P:5) (L:8) > |
| | T2 | < (B:1) (D:1) > |
| | T3 | < (C:1) (E:3) (G:7) > |
| | T4 | < (C:4) (E:3) (F:2) (G:6) > |
| | T5 | <(B:2) (D:1) > |
| 102 | T1 | < (O:3) (B:6) (P:8) (L:7) > |
| | T2 | < (B:2) (C:6) (D:7) > |
| | T3 | < (O:2) (B:3) (E:2) > |
| | T4 | < (E:3) (G:7) > |
| | T5 | < (D:1) (E:2) > |
| 103 | T1 | < (O:3) (B:2) (L:7) > |
| | T2 | < (B:4) > |
| | T3 | < (C:3) (E:4) > |
| | T4 | < (C:3) (E:5) > |
| | T5 | < (C:2) (B:2) > |

### 4.2 Efficient mining method: T-Map.

In this subsection, we describe the data mining method, namely T-Map. The input data into the T-Map is the log of mobile access patterns, which is obtained by integrating both of movement log and requested service log with time-related.

The T-Map is consisted of a header table and aggregating the access patterns into the memory in a compact form. Therefore we can mine frequent and interesting pattern efficiently. The head table are stored the frequently occurrence of mobile access patterns in the order of descending sequence access patterns. Compared with breadth-first algorithms such as Apriori [4] and its variants, which may need as many database scan as the length of the longest pattern. However, the T-Map method only needs one physical database scan when the constructing of header table. The database scan is to find all frequent mobile access patterns. Then, the frequent mobile access pattern is inserted into the header table in decreasing order of their sequence access pattern. Fig. 3 shows an example of the constructing header table. Let the minimum support count be 2. We gain the location information O, B, P and L for timestamp T1. Note that, E and F are omitted in the header table from timestamps T1. Due to the counting of E and F is infrequent. The header table is connected T-Map.

Table 3. The header table of T-Map-tree

| Location Infomation | support | Link |
|---|---|---|
| O | 4 | |
| B | 6 | |
| P | 5 | |
| L | 5 | |

After the scan of the database, the set of frequent access patterns is sorted in the order of descending sequence data. The T-Map structure is then constructed as follows: A node tale in T-Map is maintained to record the first-occurrence address and total count for distinct labels. First, create the root of the tree, labeled with "null or {}". The access patterns in each integration log data sets are processed in header table order and a branch is created for each integration log data sets. For example, the scan of the first integration log data set, "UID 100: O, B, P, L", which contains 1, 3, 5, 9 such as requested services, leads to the construction of the first branch of the tree with four nodes: <(O:1) (B:2) (P:5) (L:8)>, where (O:1) is linked as a child of the root, (B:2) is linked to (O:1), (P:5) is linked to (B:2) and (L:8) is linked to (P:5). The second T1's access patterns in integration log data sets, UID 101, contains the access patterns <(O:1) (B:3) (P:5) (L:8)>, which would result in a branch where (O:1) is linked to the root and (B:3) is linked to (O:1). However, this branch would share a common prefix, (O:1), with the existing path for UID 100. Thus create a new nod, (B:3), which linked as a child of (O:1). To facilitate tree traversal, an access pattern header table is built so that each item points to its occurrences in the tree via a chain of node-links. The tree obtained after scanning all of the integration according to timestamps is shown in Fig. 3 with the associated transformed to that of mining the T-Map.

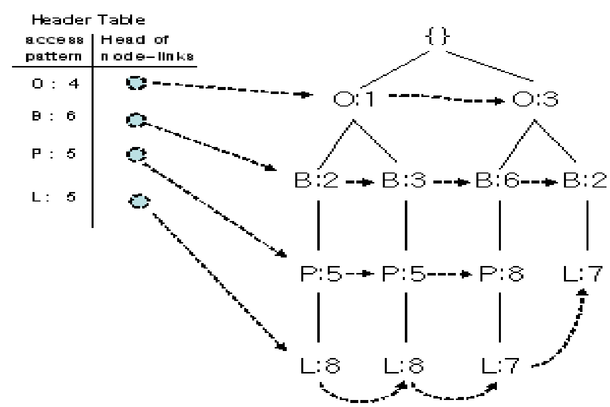


Fig. 3. T-Map for the integrated log data sets.

The T-Map can be mining frequent access pattern with time-related. The (L:8) is linked to (P:5) frequently, minimum support count is 2, and (B:2) is also frequent. However, the (B:3) is infrequent. Even if the (P:5) and (L:8) are frequent, since the occurrence of the ancestor of the (P:5) and (L:8) label is less than 2.

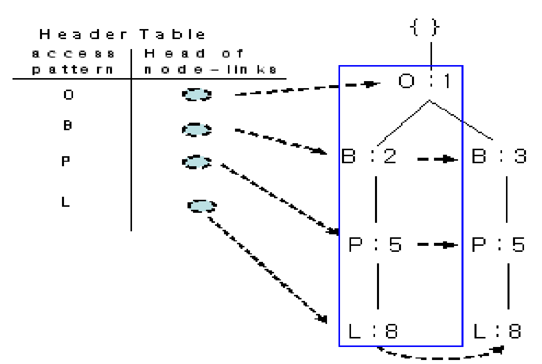


Fig. 4. Mining frequent access pattern

Therefore, the T-Map outputs the prefix access pattern <(O:1) (B:2) (P:5) (L:8)> with timestamp T1 of current T-Map and returns. Fig. 4 shows frequent use's mobile access pattern with timestamp T1 between 06:01 and 10:00.

## 5. Conclusion

In this paper, we proposed a novel data mining method, namely temporal mobile access patterns that could efficiently discover mobile user's temporal behavior patterns associated with location and requested services. Furthermore, we presented a novel data structure with temporal mobile access patterns. The advantage of a novel data structure has been compactly stored the use's behavior pattern according to location and service information in memory. Even though the information data sets consume much memory when the information data sets are storing, they are still the fastest method and consume less memory than other methods due to using the compact data structure form, T-Map, into the memory. Our technique works especially well for context-awareness data sets in mobile agent systems. Moreover, our technique may apply the health care system, mobile web system, match-making and decision support system, etc.

For the future work, we will apply T-Map more real datasets and evaluate its performance. Furthermore, we will consider that constructing of T-Map with temporal interval.

### Acknowledgments

This research is supported by the ubiquitous Computing and Network (UCN) Project, the Ministry of Information and Communication (MIC) 21st Century Frontier R&D Program in Korea.

## References

[1] C.Y Chang, M.S Chen, “Integrating web caching and web prefetching in client-side proxies”, in: Proceedings of the ACM 11th International conference Information and knowledge Management, 2002.

[2] Dan Lim, “Ubiquitous mobile Computing: UMC’s model and success”, Educational Technology & society 2(4) 1999 ISSN 1436-4522.

[3] J. F. Roddick, K. Hornsby and M. Spiliopoulou, Temporal Spatial and Spatio-Temporal data mining and knowledge discovery research Bibliography, http://www.cs.flinders.edu.au2000.

[4] J.F.Roddick and M. Spiliopoulou, “Temporal data mining: survey and issues,” Research Report ACRC-99-007, University of south Australia, 1999.

[5] J.L.Chen, Resource allocation for cellular data services using multiagent schemes, IEEE Trans. Syst. Man Cybern. 31 (6) (2001) 864-869.

[6] Jiawei Han and Micheline Kamber, Data Mining concepts and techniques, www.mkp.com or www.books.elsevier.com.

[7] J. Paik, R. Dong and M. Kim. “XML-Encoded Context Data Mining for Ubiquitous Middleware Platforms.” GESTS international Transactions on Computer Science and Engineering.

[8] P.J.Brockwell and R.A. Davis. “Introduction to Time Series and Forecasting (2nd ed.),” springer. 2002.

[9] R. Agrawal and R. Srikant, “Mining sequential Patterns,” 11th International Conference on Data Engineering, Taipei, Taiwan, Mar. 1995.

[10] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules,” the VLDB conference, Santiago, chile, September 1994.

[11] R. Srikant and R. agrawal, “Mining sequential patterns: generalizations and performance improvements,” In Proc. International Conference on Extending Database Technology, Avignon, France, Springer-verlag, 1996.

[12] Vincent S. Tseng, Kawuu W.Lin, “Efficient Mining and prediction of user behavior patterns in mobile web systems”, Information and software technology, 2005

## Biography

**SeungCheol Lee**
Education & Work experience:
B.S.: Computer Science, Cheonan Univ., 2006.
M.S candidate: Computer Science, Sungkyunkwan Univ., 2006.

Other information: Main research areas include mobile agent systems, Temporal mining, Time-series mining and XML mining. For User’s behavior pattern mining, he is interested in mining frequent user’s behavior patterns with time-related from mobile agent systems.

**Juryon Paik**
Education & Work experience:
B.E.: Information Engineering, Sungkyunkwan Univ., 1997.
M.E.: Computer Engineering, Sungkyunkwan Univ., 2005.
Ph.D candidate: Dept. of Computer Engineering

Other information: Main research areas include XML mining, XML data replication and ontology building. For XML mining, she is interested in mining frequent tree patterns, mining association rules from XML documents, and similarity computation between XML documents

**Jiwoong Ok**
Education & Work experience:
B.S.: Computer Science, Sungkyunkwan Univ., 2006.
M.S candidate: Computer Science, Sungkyunkwan Univ., 2006.

Other information: Main research areas include XML mining, XML data replication and Temporal mining. For XML mining, he is interested in mining frequent tree patterns, mining association rule from XML documents with time-series, and mining time-series between transaction database

**InSang Song**
Education & Work experience:
B.S.: Computer Science, Paichai Univ., 2006.
M.S candidate: Computer Science, Sungkyunkwan Univ., 2006.

Other information: Main research areas include XML mining, XML data replication and ontology building. For XML mining, he is interested in mining frequent tree patterns, mining association rule from XML documents, and similarity computation between XML documents

**Ung Mo Kim**
Education & Work experience:
B.S.: Mathematics, Sungkyunkwan Univ., 1981
M.S.: Computer Science, Old Dominion Univ., 1986.
Ph.D.: Computer Science, Northwestern Univ., 1990.

Other information: His main research areas are XML mining, Web databases, Database security, Data warehousing, and GIS.