International Journal of Knowledge-based and Intelligent Engineering Systems 11 (2007) 313–320
IOS Press

KeyGraph-based chance discovery for mobile contents management system

Kyung-Joong Kim, Myung-Chul Jung and Sung-Bae Cho
Department of Computer Science, Yonsei University, 134 Shinchon-dong, Sudaemmon-ku, Seoul, 120-749, South Korea

Abstract. Chance discovery provides a way to find rare but very important events for future decision making. It can be applied to stock market prediction, earthquake alarm, intrusion detection and social community evolution modeling. Similarly, contents management procedure can be regarded as a new application of chance discovery. By combining context information with contents, user's trivial behavior at a specific area can result in sudden photo creation, long-time mp3 listening, and going to specific area for content creation. Using chance discovery, such novel situation can be modeled and predictable. Recently mobile devices are regarded as a content storage with their functions such as camera, camcorder, and music player. It creates massive new data and downloads contents from desktop or wireless internet. Because of the massive size of digital contents in the mobile devices, user feels difficulty to recall or find information from the personal storage. We propose a KeyGraph-based reorganization method of mobile device storage for better accessibility to the data based on chance discovery. It can help user not only find useful information from the storage but also refresh his/her memory by using the summary of novel events. User can recall his/her memory from the contents and contexts. Using artificially generated logs from a pre-defined scenario, the proposed method is tested and analyzed to check the possibility.

Keywords: Chance discovery, KeyGraph, mobile log, smartphone, contents management system

1. Introduction

Chance discovery is to recognize a chance which is a very rare event, but with significant impact on decision making or future change [1,2]. It gives not only an awareness of chances but also an explanation about chances. This approach has been applied to various applications domains such as predicting earthquake [3], discovering new topics from WWW [4], and identifying intrusions for computer security [5].

User's contents creation and usage procedure is extremely difficulty to predict even when the size of contents continuously grows. User's trivial habit can lead to many interesting huge results such as browsing photos for many hours, going to specific areas for buying or consuming contents, and sending SMS and Calls for friends to communicate about contents. In a perspective of chance discovery, user's rare but significantly important behavior or patterns can be usefully exploited to organize and prepare information.

Personal information management is one of the hottest issues because huge number of sensors is available at this moment and they can collect all the information about users [6]. Everything about users including photo, e-mail, movie clip, computer usage, TV watching, and contexts can be stored in a unified manner [7]. Usually, data mining approach is used to extract regular or statistically significant patterns from them to predict user's information access behavior. However, there are many unexpected situation that makes user behave differently and generates novel results. For example, user forgets an exact keyword and he feels difficulty to guess the exact name. It results in huge time loss. However, it is not frequently occur. Chance discovery can be used for the situation. Also, organizing user's daily information based on the novel situation, user can easily remember daily event and understand its effect. The averaged and regular mined information can be useful to understand daily life in a bird eye's view. However, sometimes, user wants to view the daily life in the perspective of a novelty and its related situations.

ISSN 1327-2314/07/$17.00

Though it is not easy to collect everything about users, relatively easy method such as using personal mobile devices can be a partial solution for the problem. Advances in mobile computing devices have led to digital convergence. Recent mobile phones provide many functions such as MP3, camera, game, PIMS (Personal Information Management System), and so on. Using logging software [8], the user's interaction data on the phone can be stored in the inside of device or remote server and retrieved for future use. Such information can be used to enhance user's access to the contents on the phone and expand limited human's memory.

KeyGraph is one of the most frequently used methods for chance discovery [9]. Originally, it is proposed to index terms in a set of document and its purpose is to find the main point of the documents, not frequent terms. Because the KeyGraph is only based on the information within documents, it does not rely on the domain-specific corpus. If we can expand the meaning of sentence and documents to the more general one, it is possible to detect chance in many practical areas.

The purpose of this research is applying the chance discovery method to the organization of information stored in the personal database. Because the size of information is huge, efficient organization of information is critical to find information quickly and accurately. Organizing the information in a chance-based manner is similar to the mechanism of human's memory structure. Studies on human memory support the assertion that people use special landmarks for recall and the memory is organized by episodes of significant events [10]. Chance discovery algorithm is used to find landmarks among daily events and the whole memory is reorganized by episodes with landmarks identified. The landmark is regarded as a novel event for the day. Figure 1 shows episodic memory of personal databases.

In this paper, we propose a novel method to manage user's information in smartphone. User's logs such as call logs, SMS logs, multimedia logs, GPS logs, and so on are recorded and combined into an integrated log format. The logs follow the prearranged procedure to find the key events and the relationship of each pair of events applying KeyGraph algorithm which detects rare but very important event based on event sequences and major changes of event environment. Because the key is unique and memorable thing, it can be used as landmarks. Each cluster is regarded as an episode. A user can explore the personal databases using the provided landmark identifiers.

There are many alternative multidimensional data visualization methods such as correspondence analysis, multidimensional data scaling, and simple co-occurrence network. However, they deal with only global trends of the data and it is difficult to capture rare and important events. Similar to the mechanism of human brain, it is better to use rare important events to find something forgotten. If user does not know much about the contents that he wants to find, the keys can be used as a guideline for the searching.

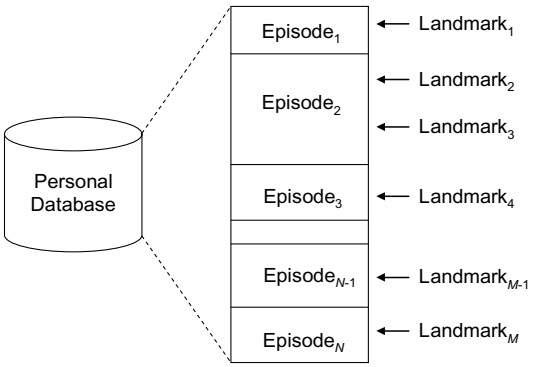

Fig. 1. Episodic memory of personal database.

2. Related works

Although there are many initial-stage papers about personal information management on mobile device, interpretation of daily life, and detection of memory landmark, they are not integrated into one system for useful applications. In this paper, we will design an integrated system of landmark detection, and information access.

ContextPhone is a context logging software for Nokia 60 series Smartphone [8] and its source is available to public. It collects information including photo, sound, battery level, location, SMS, MMS, call logs, Bluetooth, and active applications. However, it is not easy to use such software in general because Nokia 60 series Smartphone is not available in some places.

Miikkulainen proposes an episodic memory that has functions of classifying, storing and retrieving user's memory recorded in scripts using hierarchical SOM (Self-organizing Feature Maps) [11]. Hierarchical structure enables the system to reduce time for recalling the memory.

Eric Horvitz et al. attempt to re-organize personal information storage in desk top PC into an episodic style memory [10]. He learns Bayesian networks to detect landmark event from the data stored in outlook scheduler. Given schedule information, the Bayesian net-

work provides the probability of landmarks of events. Bayesian networks are used for chance discovery and it can easily deal with uncertainty in the data [12]. However, Bayesian network requires relatively high computational overhead compared to other simple models, which is the main reason of difficulty in the use of the model in computationally poor environments.

Eric Horvitz et al. also develop Bayesphone that uses client-server communication for Bayesian network whose inference is done on server-side and the results are transmitted to device through network [13]. It causes high communication cost and the device must be always online. SMILE (Structural Modeling, Inference, and Learning Engine) is a Bayesian network library for mobile device [14]. Though it supports a way to implement Bayesian network inference in mobile device easily, it cannot handle inference of large Bayesian networks. KeyGraph is relatively simple model for predicting landmark and it provides a natural view on the episodic memory because each cluster can be regarded as an episode.

Useful services using the well-organized personal information are very important for the success of personal information management. E. Horvitz et al. develop LifeBrowser and MemoryLens for more efficient access for the information on desktop [15,16]. They exploit the landmark probability inferred from the learned Bayesian networks to visualize the structure of information (stored desktop files).

MIT reality mining group develops serendipity service using the ContextPhone software [17]. The group collaborates with MIT common sense reasoning group to generate diary automatically. Because the research is at early stage, there is no concrete result about that. Only visualization tool for collected log is available in their paper. However, their work shows a new way to generate more interpretable high-level diary (interpretation) using common sense. Basic details about the common sense knowledge can be found in [18].

Recently, Sumi et al. develop a ubiquitous system to summarize user's experience on conference tour in a video but his work depends on too many extra sensors and devices to do that [19].

3. Contents management using KeyGraph

The proposed method consists of log collection, log integration, KeyGraph generation, KeyGraph analysis and information search. Figure 2 shows the overview of the system.

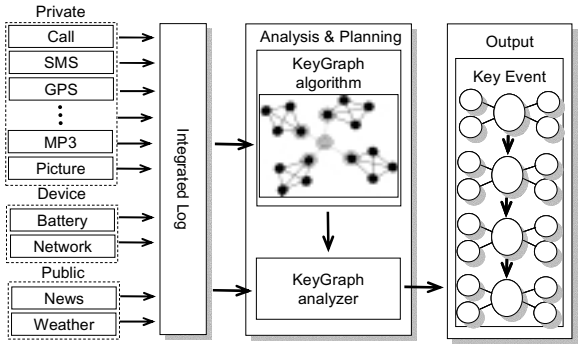

Fig. 2. System overview of the proposed method.

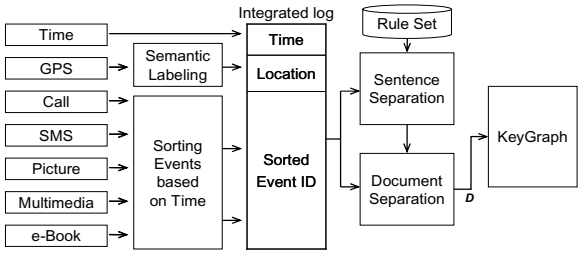

Fig. 3. Details of log preprocessing and KeyGraph generation.

Log collection module continuously gathers application usage, call & SMS, location, device status, and public web information. The logs for the several sources are sorted based on the time. The sorted log is called as integrated log. Then, KeyGraph is generated from the new log and key events are extracted. The whole memory is reorganized based on the keys and clusters. By clicking the key, user can access to the related information easily.

3.1. Log preprocessing

Time, GPS, call, SMS, photo, mp3, e-book, device status and web information can be logged into personal store. Figure 3 shows log preprocessing procedure in detail. Because GPS information is just a pair of latitude and longitude, it is required to convert them into semantic label such as name of building, street and landmark place. Pre-stored mapping table is used to do that. Other log data except location are sorted based on the time and each event is labeled as a unique ID. The naming is the combination of log type and event ID. If the event is for calling and it is the 5^{th} event of the day, the ID is Call5. Figure 4 shows an example of the naming and sorting.

Table 1
A set of rules

Log	Sentence separation rules
Time	$T < Time < T + \delta$, $T + \delta < Time < T + 2 \times \delta, \ldots$
Location	**If** (L_new≠L_old) **then** new_sentence
Call	Call events based on caller
SMS	SMS events based on sender
Mp3	MP3 events based on music title
Photo	Photo events based on location

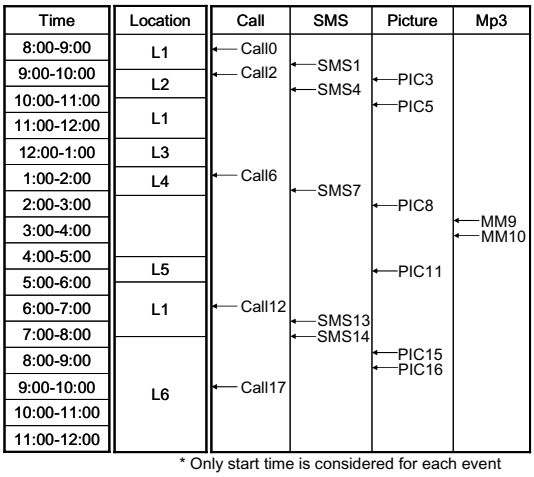

Fig. 4. An example of sorted event ID.

3.2. Document & sentence separation

Because input of the KeyGraph is a set of documents, it is required to reorganize the logs into sentences and documents. The separation of logs is done based on pre-defined rule. 24 hour data are regarded as a separate document. If the data are collected for 7 days, it means that there are 7 documents. Each document contains each day's log. Let's define the total document set as D. i represents the ith day's among N days. j represents the identifier for each event of the day.

$$D = \{d_1, d_2, \ldots, d_N\} \tag{1}$$

Each document is defined as follows.

$$d_i = \{e_{i1}, e_{i2}, \ldots, e_{iM}\} \tag{2}$$

e_{ij} means the jth event in the document. A sentence is defined as a group of events. Based on the grouping rule, there are many different ways to form the sentence. The rule set is defined as R;

$$R = \{r_1, r_2, \ldots, r_P\} \tag{3}$$

Let's define a set of sentences generated from each rule as S_{ij}. The sentence set S_i for the document d_i is defined the union of all S_{ij}.

Table 2
An example of the document

Call20# Pic2# Pic7# SMS11#.
EB24# MM12# Pic7# Call20# SMS11#.
MM5# SMS11#.
EB24# MM5#.

$$S_i = S_{i1} \cup S_{i2} \cup \cdots \cup S_{ip} \tag{4}$$

Table 1 summarizes all sentence separation rules. If time is used as a rule and the threshold is one hour, there could be 24 sentences for a day. The meaning of place is converted using predefined mapping table. It contains the mapping rule from GPS position to semantic symbols. Using open web service (http://www.naver.com), it is possible to convert it easily.

The purpose of using multiple rules is increasing the number of sentences. If the number of sentence is not enough, it is difficult to get appropriate KeyGraph. Using multiple rules, it is possible to generate many sentences although each event can be incorporated multiple times.

3.3. KeyGraph generation

KeyGraph extracts the important events and the causal structures among them from such an event sequence [9]. A document example is given in Table 2. It is composed of four sentences. Figure 5 shows an example of KeyGraph based on the document. Each cluster is composed of co-occurring frequent events. That is, events appearing frequently in D are extracted, and each pair of events that often occur in the same sentence unit is linked to each other. 'Call20, Pic7, SMS11' forms a cluster as shown in the figure. The events that are not frequent but co-occurring with multiple clusters, e.g., 'MM5' is a key event. It is rare but very important events.

The details of the KeyGraph generation are as follows. First, highly frequent events in the document are listed. Then, pairs of the events that are often co-occurred are extracted based on *local*(e_i, e_j) in Eq. (5), and their link is drawn by the solid line in Fig. 5. e means each event and g means the cluster that is com-

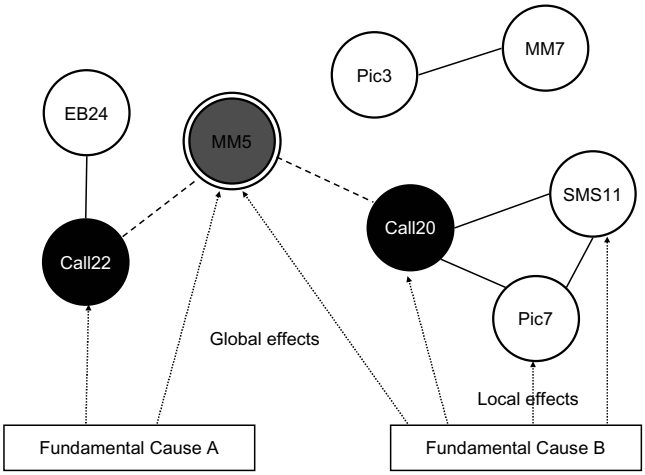

Fig. 5. An example of KeyGraph.

posed of e linked by solid lines. $|e|_s$ means the count of e in sentence S.

$$\textit{local}(e_i, e_j) = \sum_{S \in d} \min(|e_i|_S, |e_j|_S) \tag{5}$$

$\textit{global}(e_i, g)$in Eq. (6) calculates the strength between event and cluster. Event e that has the highest value summed by $\textit{global}(e_i, g)$ from every cluster is extracted as key event described as Eq. (7). Links with key events is drawn by dot lines. $|g|_s$ means the count of cluster g in sentence S.

$$\textit{global}(e_i, g) = \frac{\sum_{S \in d} |e_i|_S |g\text{-}e_i|_S}{\sum_{S \in d} \sum_{e_i \in S} |e_i|_S |g\text{-}e_i|_S} \tag{6}$$

where

$$|g\text{-}e_i|_S = \{ \begin{matrix} |g|_S\text{-}|e_i|_S & if\ w \in g \\ |g|_S & if\ not\ w \in g \end{matrix} \}$$

$$key(e_i) = 1\text{-} \prod_{\mathrm{g} \in \mathrm{G}} (1\text{-}global(e_i, g)) \tag{7}$$

3.4. Information access

Every key event arranges in order of time. User can simply search for his/her key events of a day and get access to their sub-events quickly. In short, from the keys of KeyGraph, user can access contents linked quickly and it is the purpose of this system's service. Figure 6 shows an example. If "game" is selected from the key events, game in the mobile device previously played at event 2 is executed. Similarly, e-book contents and weather information that are access together with the events are loaded.

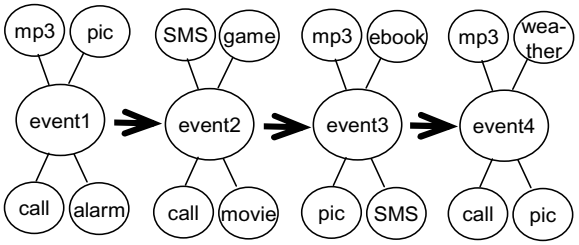

Fig. 6. An example of service generation.

4. Scenario-based analysis

The target of a scenario is daily life of an undergraduate student with Smartphone and it is summarized in Fig. 7. The logging software of the Smartphone records the events when user's activity changes. When the day's schedule finishes, contents management software lets user know the key events and the sub-events based on their co-occurrence analysis. Figure 8 shows input logs for the scenario.

Figure 9 shows a part of KeyGraph based on the user's events of the scenario. The black nodes denote key events of the graph and the white nodes denote highly frequent events. "Pic15," "Pic16," and "Pic17" mean the pictures taken by Smartphone with his friends in cafeteria. "SMS20" denotes friend's SMS to notify the important meeting with club friends. "MM34," "MM80," and "MM81" are impressive songs to the user. "Call65" means the events that user call to his girlfriend when the meeting is over. At this time, "Call65" is strongly linked to MP3 events because he goes home listening MP3.

Figure 10 is a part of the results analyzing KeyGraph. The key events are extracted on time axis,

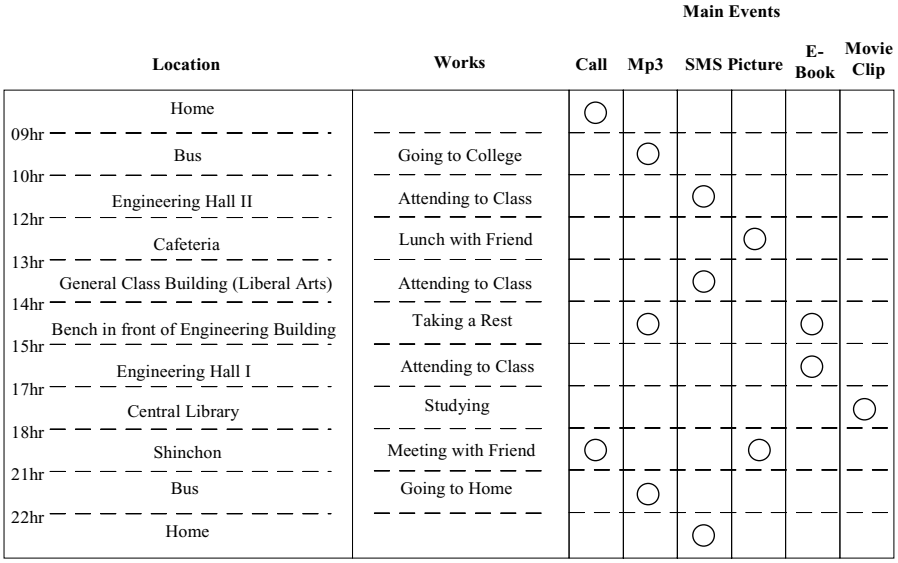

Fig. 7. User's daily life

Call

ID	Time	Name	Group	Type	Calling Time	Description
Call0	08:35:40	Ji Hye Kim	Girl Friend	Receive	5:30	Call from girl friend about going to school
Call4	09:22:12	Sung Wook Yang	Friend	Call	4:30	Call for lunch
Call21	13:51:23	Min Ho Cho	Friend	Receive	4:20	Call from high school friend

SMS

ID	Time	Name	Group	Type	Description
SMS20	13:34:23	Jae Hong Kim	Friend	Get	Please remember today's meeting!
SMS42	15:23:34	Ji Hye Kim	Girl Friend	Send	Study hard! Don't sleep!
SMS44	15:35:20	Sung Wook Yang	Friend	Get	Nice to meet you!

Multi Media

ID	Time	Title	Type
MM1, MM2, MM3, MM5, MM6, MM7, MM8, MM9, MM10, MM12	09:11:06~09:52:23	MP3track-1,2,3,4,11,12,5,6,7,8	MP3
MM45	17:01:11~17:06:02	CNN News 1	Movie
MM47	17:06:04~17:11:30	CNN News 2	Movie

Picture

ID	Time	Place
Pic16	12:13:11	Cafeteria
Pic17	12:55:00	Cafeteria
Pic54	18:30:10	Shinchon
Pic55	18:32:55	Shinchon

e-Book

ID	Time	Book Name
EB22	14:01:06~14:55:11	EB_1
EB43	15:30:06~15:48:00	EB_2

Fig. 8. Input logs artificially generated.

and the sub-events linked with the key events are displayed. The merit of the system is to allow user to access the multimedia contents easily from the keys. Instead of exhaustively checking all log information, user can access contents quickly from the key events. Instead of investigating all log information, user can check "Pic5," "Pic17," "SMS20," "MM34," "MM45," "MM48," "MM49," "Call65," and so on. If user remembers that "MM48" is closely related to something to find, he can check related contents ("Call53," "MM50," "MM52," and "MM46"). In this way, user can save the time for searching information.

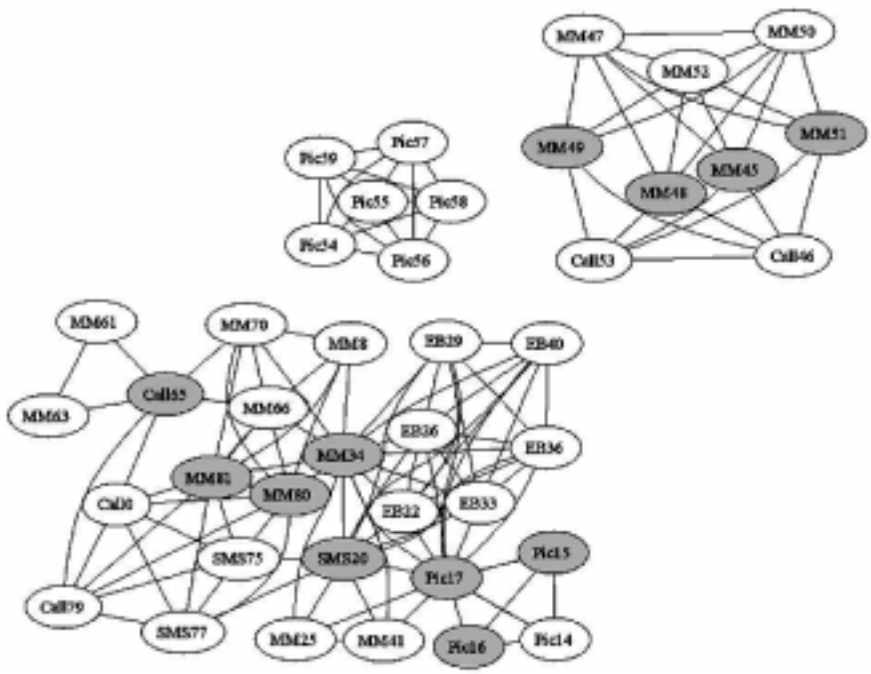

Fig. 9. KeyGraph from the scenario (MM = Multimedia, EB = E-Book, Pic = Picture).

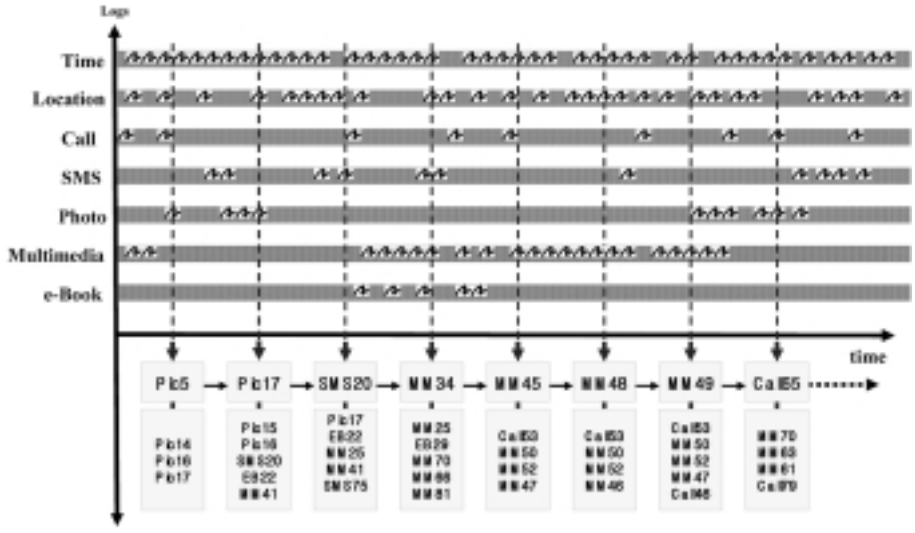

Fig. 10. The key events and the sub-events based on user's logs in time order.

5. Conclusion and future work

In this paper, we have proposed a method that can find rare but very important events based on event sequences from Smartphone using KeyGraph. We define available logs in Smartphone and build the integrated logs and transform them into documents using sentence generation rules. Analyzing the results of KeyGraph, user's information is arranged by key events. The usefulness of the system is evaluated based on scenario and artificially generated data. KeyGraph results show the possibility of the proposed method. As a future work, real logging software needs to be used for collecting user's data. Based on them, the system can be evaluated and tested in a systematic way.

Acknowledgements

This research was supported by the MIC (Ministry of Information and Communication), Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Assessment), IITA-2007-(C1090-0701-0046).

References

[1] Y. Ohsawa, Chance discoveries for making decisions on complex real world, *New Generation Computing* **20**(2) (2002), 143–164.
[2] Y. Ohsawa and P. McBurney, *Chance Discovery*, Springer, 2003.
[3] Y. Ohsawa, KeyGraph as risk explorer from earthquake sequence, *Journal of Contingencies and Crisis Management* **10**(3) (2002), 119–128.
[4] Y. Ohsawa, H. Soma, Y. Matsuo, N. Matsumura and M. Usui, Featuring web communities based on word co-occurrence structure of communications, *Proceedings of the 11th International Conference on World Wide Web*, 2002, 736–742.
[5] J.-M. Koo and S.-B. Cho, Interpreting chance for computer security by viterbi algorithm with edit distance, *New Mathematics and Natural Computation* **1**(3) (2005), 421–433.
[6] J. Teevan, W. Jones and B.B. Bederson, Personal information management, *Communications of the ACM* **49**(1) 40–43.
[7] J. Gemmell, G. Bell and R. Lueder, MyLifeBits: A personal database for everything, *Communications of the ACM* **49**(1) (2006), 88–95.
[8] M. Raento, A. Oulasvirta, R. Petit and H. Toivonen, ContextPhone: A prototyping platform for context-aware mobile applications, *IEEE Pervasive Computing* **4**(2) (2005), 51–59.
[9] Y. Ohsawa, N.E. Benson and M. Yachida, KeyGraph: Automatic indexing by co-occurrence graph based on building construction metaphor, *Proc. Of Advanced Digital Library Conference* (*IEEE ADL'98*), 1998, 12–18.
[10] E. Horvitz, S. Dumais and P. Koch, Learning predictive models of memory landmarks, 26th *Annual Meeting of the Cognitive Science Society*, 2004.
[11] R. Miikkulainen, Script recognition with hierarchical feature maps, *Connection Science* **2** (1990), 83–101.
[12] K.-J. Kim and S.-B. Cho, Uncertainty reasoning and chance discovery, *Chance Discovery in Real World Decisions*, Springer, 2006.
[13] E. Horvitz, P. Koch, R. Sarin, J. Apacible and M. Subramani, Bayesphone: Precomputation of context-sensitive polices for inqury and action in mobile devices, *User Modeling* (2005), 251–260.
[14] GeNIe & SMILE, http://genie.sis.pitt.edu.
[15] M. Ringel, E. Cutrell, S. Dumais and E. Horvitz, Milestones in time: The value of landmarks in retrieving information from personal stores, *Proceedings of Interact 2003, the Ninth IFIP TC13 International Conference on HCI*, 2003, 228–235.
[16] E. Cutrell, S.T. Dumais and J. Teevan, Searching to eliminate personal information management, *Communications of the ACM* **49**(1) (2006), 58–64.
[17] N. Eagle, *Machine Perception and Learning of Complex Social Systems*, Ph.D. Thesis, Program in Media Arts and Sciences, Massachusetts Institute of Technology, 2005.
[18] P. Singh, B. Barry and H. Liu, Teaching machines about everyday life, *BT Technology Journal* **22**(4) (2004), 227–240.
[19] Y. Sumi, S. Ito, T. Matsuguchi, S. Fels and K. Mase, Collaborative capturing and interpretation of interactions, *Pervasive 2004 Workshop on Memory and Sharing of Experiences*, 2004, 1–7.