Online ISSN 1848-3380, Print ISSN 0005-1144
ATKAFF 54(4), 401–414(2013)

Keedong Yoo

Capture Knowledge on the Spot: Toward the Autonomous and Pervasive Service of Context-Rich Knowledge

DOI 10.7305/automatika.54-4.411
UDK 004.896:004.855.6
IFAC 2.8.1; 4.0.5

Original scientific paper

Knowledge must be acquired not only at the moment when it is presented, but also at the site where it is applied to. To guarantee the immediate acquisition of context-rich knowledge at anytime and anywhere, fully-automated as well as pervasive capabilities must be considered together. This paper proposes a methodology to capture knowledge on the spot in an autonomous and pervasive manner by deploying the Smartphone as a sensor to monitor and gather dialogue-based knowledge and context data. Smart-ConKAS (SMARTphone-based CONtextual Knowledge Acquisition System), a prototype system, is implemented to validate the proposed concepts.

Key words: Automated knowledge acquisition, Pervasive computing, Autonomous computing, Cloud computing, Smartphone, Knowledge management

Uhvatite znanje na licu mjesta: ususret autonomnoj i prožimajućoj usluzi sadržajnog znanja. Znanje se stječe ne samo u trenutku kada je predstavljeno nego i na mjestu gdje se primjenjuje. Kako bi se jamčilo trenutno stjecanje znanja bilo kada i bilo gdje moraju se uzeti u obzir potpuno automatizirane i prožimajuće sposobnosti. U ovom radu predložena je metoda stjecanja znanja na licu mjesta na autonoman i prožimajuć način korištenjem pametnog telefona kao senzora za nadgledanje i skupljanje znanja i podataka. Smart - ConKAS (SMARTphone-based CONtextual Knowledge Acquisition System) je prototip koji je korišten kako bi se potvrdio predloženi koncept.

Ključne riječi: automatizirano stjecanje znanja, prožimajuće računarstvo, autonomno računarstvo, računarstvo u oblaku, pametni telefon, upravljanje znanjem

1 INTRODUCTION

A variety of methodologies to automatically capture knowledge have been proposed, however each of them was reported to have certain level of limitations not only in the degree of automation but also in the quality of captured knowledge. The efficiency of the automated knowledge acquisition is still unsatisfactory due to over-complex algorithms and immature methodology [21]. Additional manual operation must be inevitably performed, because the automated acquisition cannot effectively, namely appropriately, gather knowledge which resides in the human brain [6]. The quality of acquired knowledge also disappoints knowledge users, because lack of context information which illustrates when and where the given knowledge is able to be applied makes them degrade the applicability of knowledge [11]. Knowledge must be captured at the moment when it is exhibited as well as at the site where it is applied to simultaneously acquire its context data [22]. To guarantee the immediate acquisition of context-rich knowledge at anytime and anywhere, fully-automated as well as pervasive capabilities must be additionally considered: the border line between this study and the conventional ones.

To capture knowledge on the spot with its related context data, the subject who possesses and utilizes knowledge must be monitored continuously, nearly around the clock. A knowledge worker, or a knowledge possessor, who has the knowledge to be captured performs his or her jobs using special, and so preserving-worthy, knowledge. His or her dialogues, writings, and activities exhibited during performing jobs are the live sources to be managed with care. Applying proper technologies to each source, fully-automated, and so autonomous, as well as pervasive acquisition of knowledge can be achieved [23, 24]. The context data which plays the role of the meta-knowledge can also be captured by monitoring the knowledge worker. The context data can be defined as any information characterizing the situation of a task session or interaction between a user and his or her service world [25]. The meta-

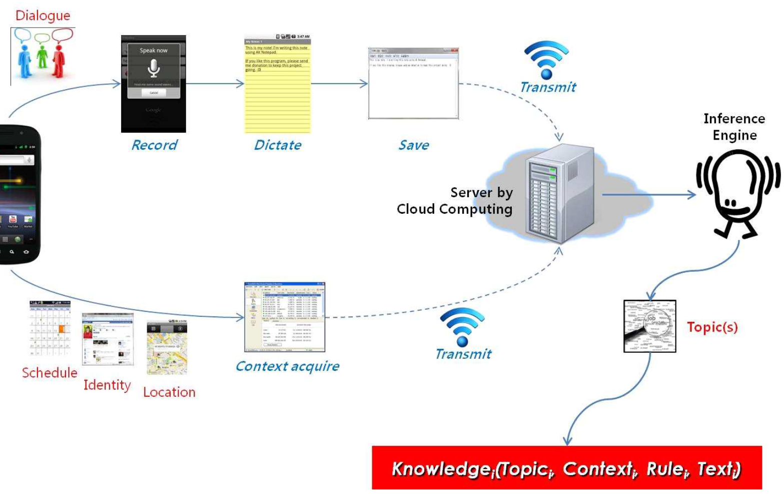

Fig. 1. Conceptual Framework

knowledge-based context data can be understood as the context data which explains the situation of the knowledge worker who is utilizing the given knowledge, therefore it is a kind of the user context.

User context can be summarized into three categories [18]: information of the user (e.g. knowledge of habits and emotional state), the user's social environment (e.g. co-location of others, social interaction, and group dynamics), and the user's tasks (e.g. spontaneous activity, engaged tasks, and general goals). To characterize the given knowledge, several types of context data, such as the identity that explains who the knowledge worker is as well as which area he or she specialized in, the time and location he or she uses knowledge, the event (schedule) the given knowledge is to be applied, and the topic (keyword) the exhibited knowledge is about, etc., are combined together.

Mobile devices, such as Smartphones and Smartpads, are widely used nowadays because of their multiple functionalities based on wireless data communications and open APIs (application programming interfaces). As developers of mobile devices design devices to be lighter, smaller, and more versatile, mobile devices tend to be multi-functional enough to be a mandatory item of human life: every individual carries and uses them anytime and anywhere. Even in performing ordinary jobs, individuals carry one or more mobile devices in their hands or pockets, therefore mobile devices can be used as a sensor to monitor knowledge workers' dialogues, writings, and activities in a real-time basis. Once a mobile device monitors and records what a knowledge worker talks, writes, and acts, then it can also transmit them to a server whose address has been designated in advance by the user. Using the paradigm of cloud computing, a networking model for enabling convenient on-demand network access to a shared pool of configurable computing resources [16], transmitted data can be analyzed so that the meaning of each can be identified. Context data can be also captured and transmitted to a server using mobiles devices in a very convenient way, because most of context data have been preset and stored in the mobile device using a user profile and a scheduler. Figure 1 shows how a mobile device contributes to capture knowledge (dialogue-based) and context data to constitute the autonomous and pervasive acquisition of knowledge.

The objective of this paper is to propose a methodology to capture knowledge on the spot in an autonomous and pervasive manner, which deploys the Smartphone as a sensor to monitor and gather dialogue-based knowledge and its context data. To demonstrate the validity of the proposed concepts, Smart-ConKAS (SMARTphone-based CONtextual Knowledge Acquisition System), a prototype system, is also implemented.

2 RELATED WORKS

2.1 Context-based knowledge management

Thinking knowledge with considering context has not been matured comparing to other context-related researches. Since context can give guidance about when and why a piece of knowledge is used, considering context in knowledge use is very necessary to enhance the applicability of knowledge stored in a knowledge repository.

Takashiro and Takeda [19] proposed a method for acquiring and utilizing personal knowledge in computer systems. They developed a system called 'MindHeap' that helps users acquire personal knowledge by browsing WWW (World Wide Web) hypertexts. This approach uses the situational context which has important effects on acquiring and reminding knowledge. MindHeap can help users organize knowledge on 'www' pages by finding and recording related words as situational context for topics specified by users. The system extracts a list of related words as situational context, and records it according to the topics. Thus, topics and their contexts are accumulated, then they can work as knowledge for user. Users can retrieve pages that have the similar situation to the current page. This work tried to store knowledge according to the context by which the given knowledge is triggered, therefore, this work can be regarded as the early attempt to manage knowledge with its context.

Voida et al. [20] suggested the necessity of integrating virtual and physical context to support knowledge workers. As an answer to this proposition, they classified jobs of knowledge workers according to virtual and physical context, and developed a design and architecture of the 'Kimura' system. The Kimura system augments and integrates independent tools into a pervasive computing system that monitors the user's interactions with a computer, an electronic whiteboard, and a variety of connected peripheral devices and data sources. The Kimura system made it possible to use virtual context from the user's desktop actions to help classify, interpret, and visualize other forms of virtual and physical context. Also, the system enabled integrating virtual and physical context information into visualization of the user's disparate activities to help the user interpret and act on this available information. Knowledge workers' virtual context can be visualized, and therefore knowledge workers' virtual and physical context can be synthesized to identify their activity patterns.

Anerousis and Panagos [1] addressed the issue of pervasiveness using regular landline and cellular phones as pervasive devices to capture and deliver voice knowledge. The voice knowledge is the knowledge communicated verbally and its examples are messages, commentaries, conversations, and conferences. They developed a system entitled 'TotalKnowledge' to pervasively create and access knowledge in the form of the voice. The TotalKnowledge is designed to search voice assets on the basis of their metadata and content, to view and edit metadata, and to archive existing short term and long term assets. This work enabled the pervasive knowledge management by using mobile devices' capability of real time-based knowledge registering and querying.

Huang and Tao [7] presented a knowledge interoperation reference model and the concept of Context Knowledge Grid to leverage the contextual knowledge in modem enterprises and to enable the interoperation with other knowledge frameworks such as the Semantic Web and the Semantic Grid. By defining a contextual knowledge structure model, they tried to connect knowledge to context, and therefore to leverage the understandability of knowledge use. Relationship-defined context and knowledge are designed to be stored in the ontology-based repository. This work reasonably explains the relationship between context and knowledge by functionally matching them and generally deriving a model.

Nunes et al. [15] introduced a model for managing context-based knowledge, which addresses the creation, storage, and reuse of contextual knowledge, encompassing the representation, capture, storing, comparison, and presentation of knowledge in the setting where the work process activity is performed. They stressed the necessity of considering context in managing knowledge and defined an ontology-based structure for context elements. This work verified not only what kinds of context information are needed to be defined when process-based knowledge is executed, but also how context information can be described to preserve the characteristics of processes encompassing context information and knowledge.

Previous researches contribute to the context-based knowledge management, especially focused on three aspects: the relationship between knowledge and context, the role of context in understanding and utilizing knowledge, and the implemented artefacts incorporating knowledge with context. On the contrary, previous researches indicate certain level of limitations in two perspectives: the extent of implementation and automation. Most of previous researches just introduced concepts or theories without real implementation to prove the validity of the proposed ones. Although some of previous researches tried implementing prototypes, most of them lacked in automation, which is one of the required functionalities that a ubiquitous and pervasive system must have: this point keeps this research and previous ones separate.

2.2 Autonomous computing

'Autonomous Computing' concerns self-generating features of computer-based services, providing adequate computer services in a proactive manner by continuously monitoring and identifying users' context denoting their current situations. An autonomous system, therefore, not only makes decisions on its own, but also communicates with users continually in a real time basis using another high-level policies; it will continuously monitor and identify users' situations and automatically adapt its services to changing situations surrounding users. The difference between the autonomic computing proposed by IBM and the autonomous computing does exist in system's main interest: the former focuses on the system itself, while the latter concentrates on the system's outputs, namely services.

In a self-service generating autonomous system, therefore the human operator is charged with a new role also.

He or she does not need to control the system directly either. Instead he or she defines rules between context and services, and the ways of communication between users and computers. An autonomous computing-based system can have functionalities as follows:

1. Self-Awareness: Automatic detection of users' situations using context data, such as location, identity, time, schedule, etc.;
2. Self-Service generation: Automatic as well as proactive inferring of proper services based on identified users' situations;
3. Self-Learning: Automatic learning of the relationships between context and service;
4. Self-Expansion: Automatic addition of rules between context and service into knowledge and context repository.

Once the dictated dialogue and context data captured by the Smartphone are transmitted to a server operated on a cloud computing basis, further treatment to extract the topic (keyword) of the transmitted data is processed as the concept of Autonomous Computing advocates. By extracting topics of the transmitted text-based data, each document can be classified in terms of the extracted topics or keywords. The document can be deemed as knowledge itself, because it may contain various kinds of knowledge applicable to the similar situations. Because the topic of the document is the topic of knowledge, knowledge possessors' knowledge can be categorized according to the topic. Besides the topic resulted from analysing the document, knowledge possessors' context data which have been identified and transmitted from the context acquisition subsystem must be applied to conclude the relationship between the knowledge and the situation that the knowledge is deployed. Context data, such as knowledge possessors' identities, location, time, and schedule, can play the role of the meta-knowledge by uniquely depicting the knowledge used in the given situation. By associating the document and context data, knowledge can be stored in the knowledge base with the form of a business rule.

Suppose a case that 'Professor A' is discussing about how to make payment for a purchased book that is expensive but that is necessary for his research. General payment method is known as one of 'research fund', 'personal expense', and 'mileage points'. If the theme (topic) of the book is related with his research then he can use the 'research fund'; otherwise payment must be made through one of 'personal expense' and 'mileage points'. In this case, the most important information that determines the method of payment is the theme of the book. Next important information is the balance of the research fund, because either of the rest, personal expense and mileage point, must be unavoidably considered if the balance is not enough. Possible business rules for this situation are as follows:

```
Rule1:
IF Theme_of_Book = Theme_of_Research THEN Method_of_Payment = 'Research_Fund'
                    ELSE Method_of_Payment = ['Personal_Expense' | 'Mileage_Point']
Rule2:
IF Balance_of_Research_Fund >= Price_of_Book THEN Method_of_Payment = 'Research_Fund'
                    ELSE Method_of_Payment = 'Research_Fund' + ['Personal_Expense' | 'Mileage_Point']

Theme_of_Book = [Title_of_Book | Keyword_of_Book]
Theme_of_Research = [Title_of_Research | Keyword_of_Research]
Method_of_Payment = ['Research_Fund' | 'Personal_Expense' | 'Mileage_Point']
Title_of_Book = *The title of the book revealed in the cover page*
Keyword_of_Book = *The keywords of the book defined by the authors and revealed in the bibliographic page*
Title_of_Research = *The title of research as an instance of the class 'Research' stored in a database*
Keyword_of_Research = *The keywords of the research as an instance of the class 'Research' stored in a database*
Balance_of_Research_Fund = *Amount of research fund remained and stored in a database*
Price_of_Book = *The price of the book revealed in the bibliographic page*
```

Information included in these rules cannot be directly determined by simply monitoring the dialogue. The context data, such as the identity of the participant and the title of the book, which can be directly obtained by monitoring the dialogue, can constitute required information to form rules through the causal chain relationship. For example, the data of 'Professor A', the identity of the participant, can determine the title and keyword of the research by querying corresponding research he is responsible for. The title of the book also underpins the business rule, because it addresses the keywords and the price of the book. Causal chain relationships between the context data and the rule-constituting information can be identified as follows:

```
Relationship1:
'Identity' determines 'Title_of_Research' AND 'Keywork_of_Research' AND 'Balance_of_Research_Fund'
Relationship2:
'Title_of_Book' determines 'Keyword_of_Book' AND 'Price_of_Book'

Identity = "Name and ID number of the participant"
```

Meanwhile, the topic of the dialogue influences the business rule by identically depicting which rules can be alternatively related with the event included in the dialogue. The topic of the dialogue cannot be regarded as the context data, because it cannot be directly obtained by monitoring the dialogue but because can be identified by further analysis. Once the topic of the dialogue is identified, a set of knowledge can be automatically formulated by combining the topic, context data, and the text-based document containing the contents of the dialogue. If the topic of the dialogue is concluded as the 'Book Payment', then a set of knowledge can be expressed and stored as follows:

```
Identity[i] = 'Professor A'
Title_of_Book = 'XXXXXXX'
Rule[i] = 'Rule1' + 'Rule2'
Document[i] = 'Document001'
"text-based document containing the contents"
Rule1 = IF Theme_of_Book = Theme_of_Research THEN Method_of_Payment = 'Research_Fund'
                    ELSE Method_of_Payment = ['Personal_Expense' | 'Mileage_Point']
Rule2 = IF Balance_of_Research_Fund >= Price_of_Book THEN Method_of_Payment = 'Research_Fund'
                    ELSE Method_of_Payment = 'Research_Fund' + ['Personal_Expense' | 'Mileage_Point']

IF Topic[i] = 'Book_Payment' AND Context[i] = Identity[i] + Title_of_Book[i]
THEN INSERT INTO Knowledge[i] = Knowledge(Topic[i], Context[i], Rule[i], Document[i])
```

The topic of the dialogue('Topic[i]'), the context data('Context[i]'), and the text-based document('Document[i]') can be automatically identified and obtained from the sensors monitoring the dialogue and the applications dictating and analysing the texts in the document. Business rules can also be automatically triggered from a knowledge base by relating the topic of the dialogue with the context data. A piece of knowledge can be automatically defined and stored with the form of a business rule: autonomous computing-based knowledge acquisition can be accomplished. This kind of rule-based knowledge can be specified not only by manual operation during the development phase, but also by automated rule inference of the ontology-based knowledge base. Figure 2 shows the procedure to autonomously determine the topic of knowledge.

3 STRATEGY TO AUTOMATICALLY IDENTIFY THE TOPIC OF KNOWLEDGE: TEXT MINING

3.1 Support Vector Machines as the Classifier

The indexing and retrieval algorithm of information retrieval technology can be applied to automatically identify the topic of each electronic document and classify it according to predefined knowledge categories. Identifying the topic of knowledge is important in that the topic (or the keyword) indicates the subject of knowledge embedded in the document. To extract the topic of knowledge based on predefined knowledge categories, text mining techniques can be employed [12].

The topic of an electronic document is used further for storing the document in a knowledge repository. This topic can indicate the situation (or the context) of the knowledge holder and play a role as meta-knowledge that can directly explain the content of the knowledge embedded in the document. By combining the topic with other context data, a specific piece of knowledge can be identically defined. In other words, contextual knowledge can be obtained by defining knowledge according to related context data. Based on such context data which define the context of knowledge use, knowledge can be accurately selected and automatically serviced to the user if the user's context data explaining the situation that surrounds the user have been captured and concluded, that is, an autonomous process for the knowledge service, from the automated acquisition of knowledge to its proactive dissemination, can be realized.

Among various algorithms of text mining, this research deploys the Support Vector Machines (SVMs) for its adequacy of application. SVMs try to find the hyperplane

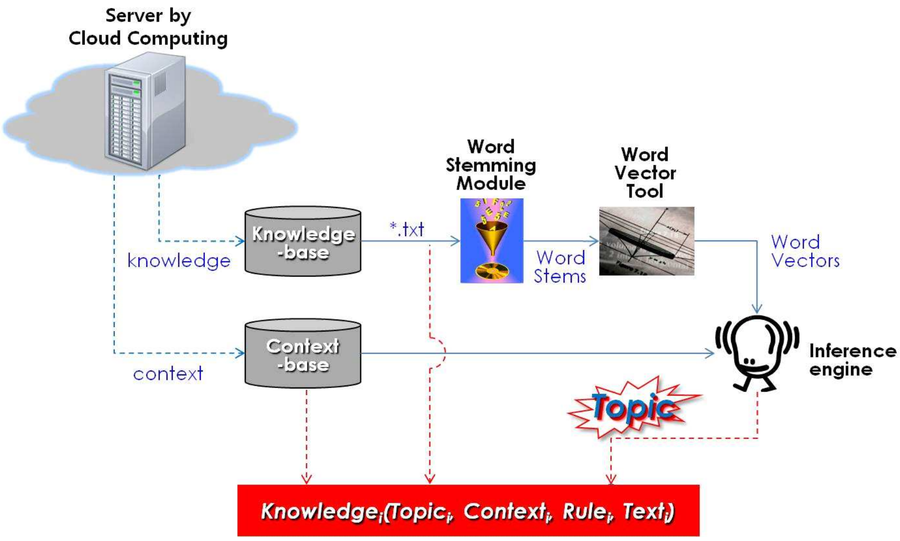

Fig. 2. Procedure to Determine the Topic (Keyword) of Knowledge

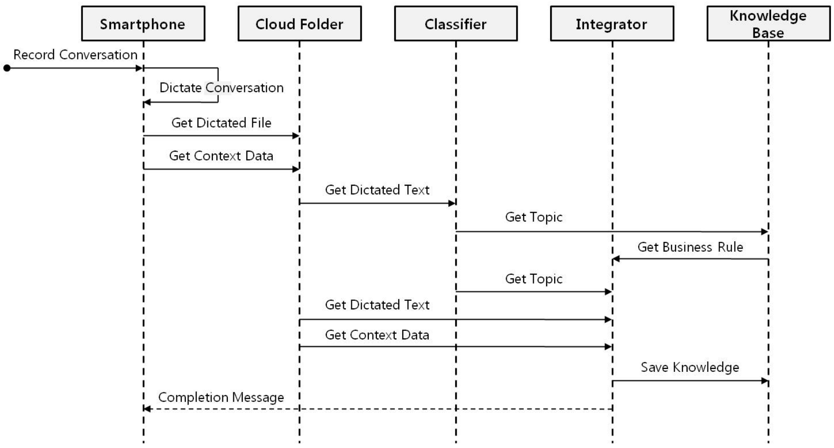

Fig. 3. Sequence Diagram of the Prototype System

which produces the greatest possible margin among the boundary points to separate positive and negative training samples. SVMs are adequately applicable to the topic (or the keyword) classification for several reasons: a high dimensional input space, few irrelevant features, sparse document vectors, and linearly separable characteristics of text categorization problems [9]. Based on these characteristics, several researches have proved SVMs outperform other machine learning algorithms with respect to efficiency as well as accuracy [13, 2, 17, 9, 5].

This research used the LibSVM v2.81 [3] to implement the SVM classification. Word stems and vectors must be provided before the LibSVM (the classifier) performs its job. Therefore the word stemming and vector creating tool of Yale [14], a free open-source environment for KDD and machine learning, was employed.

3.2 University Ontology-based Predefined Categories

To guarantee the accuracy of classification, SVMs must be trained using pre-categorized documents in advance. In most of cases, SVMs are trained using publicly well-known pre-defined categories, such as the Reuters-21578 and the Ohsumed corpus. Because these categories are very large-sized but have limited number of subjects, only selected part is used for training while the rest is used in examining the accuracy of prediction. Alternatively, predefined categories can also be provided by consulting the related ontology model which describes the objects and their relationships in the real world. In an ontology model, every object can be categorized hierarchically with the relationship of the superclass and subclass. Therefore, the ontology model of a given situation has the same struc-

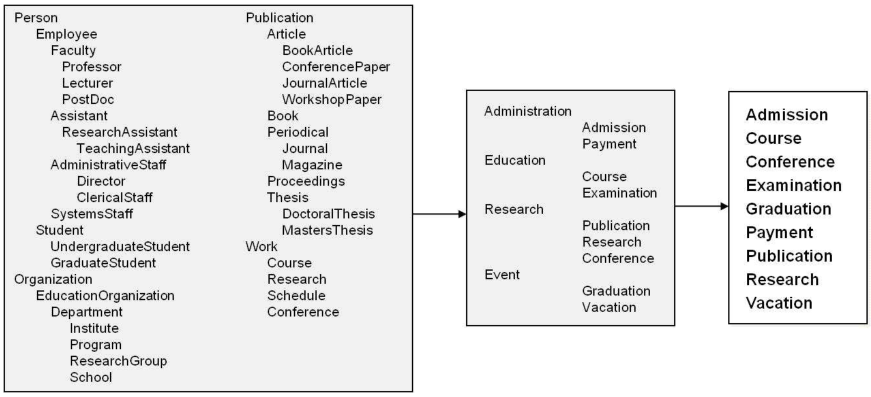

Fig. 4. University Ontology-based Predefined Categories Smart-ConKAS (SMARTphone-based CONtextual Knowledge Acquisition System)

ture with the term categories of the situation. Indeed, if the class structure of the ontology has been officially validated to be formal or standardized, the ontology model can be more preferred to conventional corpus-based categories because of ontology's advantage in concept representation.

This research adopted a set of ontology-based categories exemplifying university-based activities. To ensure the validity of category definitions, the university ontology provided by the department of computer science at the University of Maryland (http://www.cs.umd.edu/projects/plus/SHOE/onts/univ1.0.html) was employed. This ontology defines elements for describing universities and their activities and includes concepts such as departments, faculty members, students, courses, research, publications, and works. This ontology is a draft and provides a rough picture of a university, and thus, by applying WordNet v2.1 (http://wordnet.princeton.edu), slight modification was made Fig. 3: First, only event-oriented objects were selected because nonevent-based objects such as 'persons' and 'organizations' are not suitable to present which topic of conversation is being made in a dialogue. Nonevent-based objects can explain only the features of objects, that is, they cannot explain the situations or conditions that surround the dialogue participants. Thus, it is more reasonable to consider action- or event-related classes in the ontology because only context-based objects are to be considered in context categories. Second, some event-oriented classes were added, because the university ontology does not fully account for events that can be observed in real universities. Therefore, some realistic events such as 'admissions,' 'examinations,' 'graduation,' 'payments,' and 'vacations' were additionally included to better describe the university context typically experienced by university members. Finally, the superclasses and subclasses were merged into classes at the same level. As in Reuter-21578, the distinctions between levels were eliminated so that the SVM module could compare each class equally during the search for proper categories. All the modifications were based on hypernyms or holonyms (a holonym is a word that names the whole of which the word is an important part) and hyponyms (a hyponym is a subordinate word that is more specific than the given word) provided by WordNet v2.1.

To train the SVM module and prepare the prediction model, 30 sample documents for each category were collected. Each sample document was prepared by browsing webpages related to each category because each webpage is entitled by predefined and so trustworthy topic (the category). Because of the relatively small number of categories, the accuracy of training trials approached 100%, despite the small number of training sets (a total of 270 sets). Based on the predefined categories, new documents were classified.

4 SMART-CONKAS (SMARTPHONE-BASED CONTEXTUAL KNOWLEDGE ACQUISITION SYSTEM)

4.1 Overview

The Smart-ConKAS, the prototype system, enables the autonomous as well as pervasive acquisition of knowledge by using the Smartphone based on autonomous and cloud computing technologies. In this prototype system, Smartphone acts as a sensor to record conversations, the source of knowledge, in a dialogue as well as to capture context data, the meta-knowledge. The STT(Speech-to-Text) application installed in the Smartphone initiates its function of dictation by transmitting the recorded data to the vendor-providing server for speech recognition. Dictated results are again sent to the Smartphone and displayed on the screen to wait user's confirmation. Time for completing dictation can be various depending on the condition of network traffic, however in most cases dictation can be completed within a few seconds averagely. Dictated results are usually saved in text format, and they are transmitted to a pre-designated server operated on the cloud computing environment. Context data are also transmitted to the server and temporarily saved in the server. At the server, the text-based electronic document containing the contents of conversations is analysed by SVM classifier to extract its topic. Once the topic of the document, that is the topic of the dialogue, is identified, the system stores the texts in the document with respect to the topic and the context data. If any business rules related with the topic exist, then they are extracted from a knowledge base to

Table 1. Example dialogue concerning how to make a research proposal

Yoo:	Hi, how are you?
Kim:	Hmm, so so, nowadays I have been very busy with preparing a proposal for a research project.
Yoo:	What kind of research projects do you mean?
Kim:	A research project sponsored by the Microsoft research center. Have you heard about the collaborative research between a company and a university?
Yoo:	Sure I have. It's very helpful to universities to earn fund for research and technology development.
Kim:	Yes it is. But preparing a proposal is very tough work, because many applicants also try to get contracts with companies.
Yoo:	Of course it is. But once you get the contract then you can concentrate on study and research only without worrying about other stuffs, like research fund, personnel, and equipment. Every stuff required to do research can be sponsored by the contracted company.
Kim:	Then do you have any idea about an attractive proposal?
Yoo:	Let me see. First, you must concretely specify the goal or objective of your research project by listening target company's current situation. What the company aims to get from the research project or what problems the company want to solve through the collaboration with you might be one of the key considerations, I think.
Kim:	You got it. Then what about the research fund? I mean, how I can estimate the amount of fund sponsored by the company.
Yoo:	Just follow the directions provided by the company. As I know, the Microsoft estimates the amount of research fund based on the number of people who participate in the research project and the amount of equipment that are required to be set. Of course the lower the better, I mean too much amount of research fund can be doubted. Sharing research resources like people and equipment is one important reason for collaborative research project. I can give you a sample of the research fund estimation once I submitted to the Microsoft. It will be helpful to you.
Kim:	Yes please, it must be very helpful to prepare my research proposal. Thank you very much.
Yoo:	You're welcome. Good luck!

be stored together with the topic, the context data, and the text as a set of knowledge. Figure 4 shows the procedure how Smart-ConKAS acquires knowledge in an autonomous and pervasive manner.

Smart-ConKAS is developed using JDK v1.5.0_06 under Java2 runtime environment to enhance the high interoperability and the ease of implementation. It initiates knowledge acquisition by recording and dictating knowledge possessors' dialogues using STT module of Android v2.3.3 (Gingerbread) keyboard.

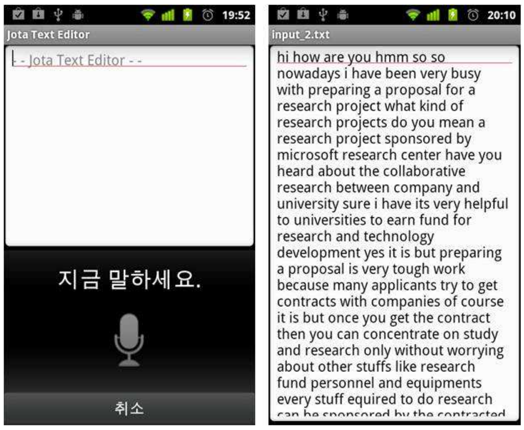

Fig. 5. Recording standby(left) and Dictation completed(right)

4.2 How Smart-ConKAS works

For the effectiveness of explaining how the prototype system works, suppose an example dialogue about a research proposal between two research associates, as shown in Table 1.

The Smart-ConKAS initiates its process by providing 'Recording Standby' signal to users, and begins recording as users start speaking. When speaking is finished or paused, dictated results are displayed and the mode of 'Recording Standby' is again activated. When the user terminates the process of recording, it automatically transmits the dictated data to the cloud folder of the designated server operated on the cloud computing environment. Figure 5 shows the recording and dictation process.

When transmitting the dictated text file to the server, it also simultaneously transmits corresponding context data which depict when and where the file has been made as well as whom the file has created by. Figure 6 shows the context data stored and obtained by the Smartphone, and Fig. 7 shows the programming codes to perform transmission.

Fig. 6. Context Data Stored in Smartphone (user profile(left), location(middle), and schedule(right))

Once the text file and context data are transmitted and saved in the cloud folder of the server, they can be synchronously accessed using a laptop or a desktop and the Smartphone. Analyzing the text file to extract the topic must be processed at the server-side, because currently the capacity of Smartphone is not enough to perform the job Fig. 8.

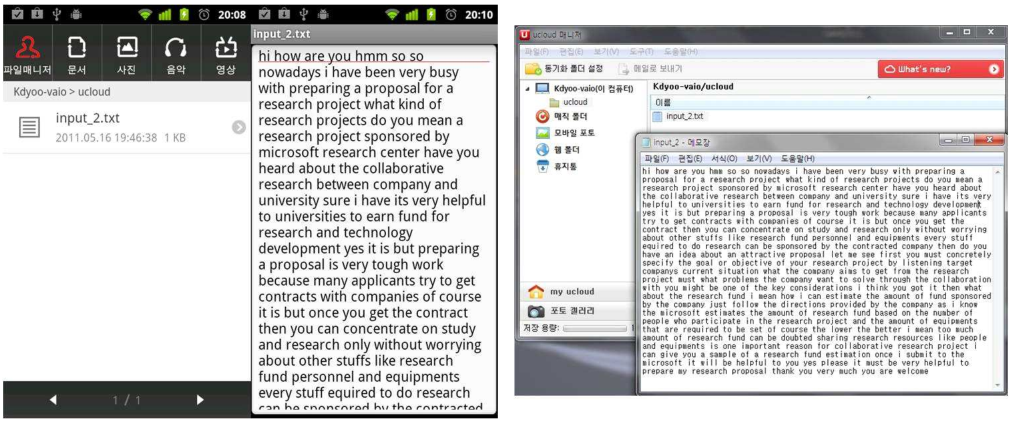

Fig. 8. View of the text file access via Smartphone(left) and laptop(right)

```
public KMclientsocket (Knowledge theKnowledge, KMContext theContext,
String KMhostname) throws IOException {

        // connnecting a KMS
        Socket KMSocket = new Socket(KMhostname, KMServerPort);

        // transmitting knowledge and its context
        ExchangeKM(theKnowledge, KMContext, KMSocket) ;

        // disconnecting
        KMSocket.close();
}

public class KMContext{
        string      ID;
        string      Location;
        String      Schedule;
}
```

Fig. 7. Programming Codes for Transmitting Dictated Results('textKnowledge') and Context Data('KMContext')

The text file stored in the server is to be moved to a laptop or a desktop to perform the topic identification by inputting it the word stemmer, the word vector tool, and the SVM-based classifier in order. The resultant value of vector of the dialogue are shown in Fig. 9.

Calculated word vector must be reformatted so that the SVM module can read it as the input. Figure 10 shows the codes for converting the format and reading word vectors.

The result from the SVM module is extracted as the number that stands for each category, which denotes the corresponding topic. There exist 9 categories in this research, and the result indicates the numbers of each category. Figure 11 is the result of classification. In the figure, '1 8.0' means 'document no.1 belongs to the eighth category', in other words, the topic of document no.1 is 'research'.

Finally, a set of knowledge must be stored in a knowledge base according to the identified topic and related context data. Since the topic plays role of the specific category to store knowledge, the context data, the document containing dictated conversations, and business rules (if any) must be moved under the topic. Figure 12 shows the programming codes to move the document under the resultant topic category. After completing storing the document under the destination category, the input document temporarily stored in the cloud folder needs to be deleted for the efficient management of the server.

4.3 Performance

The performance of SMART-ConKAS needs to be evaluated with respect to two factors: the extent of automation and the accuracy of acquisition. To examine the performance with considering these factors, the focus group interview as the qualitative approach and the statistical analysis as the quantitative approach were performed.

4.3.1 Focus Group Interview to Measure the Extent of Automation

The primary purpose of this research is to fully automate the process of knowledge acquisition. Therefore examining the extent of automation in the entire process for acquiring knowledge is necessary. The extent of automation can be regarded as a qualitative measure, because the extent can be decided by users and because it can be differently felt from user to user. Of course it is clearly observed that the entire process of knowledge acquisition has been automated through the prototype system. However, this fact cannot sufficiently explain how much a user satisfies and trusts the automated function of the prototype system. Therefore, a 10-member group of experts was organized and questioned.

At first, each member was asked to experience the prototype system by loudly reading a script selected from a newspaper article. Most of members repeated the experiment until they had examined the functions and identified the features. After finishing the experiment, at second, each of them was asked to rate the level of automation using the 7-point Likert scales, from 'none-automated' to 'fully-automated'. Based on the marked point, finally, each member was asked to explain the reason why he or she determined the point by specifying features that are fully-, semi-, and none-automated. Table 1 shows the results from the focus group interview.

Most of members of the focus group evaluated the prototype system highly automated the process of knowledge acquisition, as

nowadai	heard	applic	attract	want	particip
busi	collabor	contract	concret	solv	lower
prepar	compani	cours	specifi	consider	doubt
propos	univers	concentr	goal	think	share
research	help	studi	object	estim	resourc
project	earn	worri	listen	amount	reason
kind	fund	stuff	target	follow	give
mean	technolog	personnel	current	direct	sampl
sponsor	develop	equip	situat	know base	submit
microsoft	tough	requir	aim	Number	pleas
center	work	idea	problem	peopl	thank
					welcom

1:-1 2:-0.839486 3:-1 4:-1 5:-1 6:-1 7:0.993548 8:-1

Fig. 9. Resultant word stems(above) and vector(below)

```
if(predict_probability == 1)
{
        if(svm_type == svm_parameter.EPSILON_SVR || svm_type == svm_parameter.NU_SVR)
        {
                System.out.print("Prob. model for test data: target value = predicted value + z,\nz: Laplace distribution
                e^(-|z|/sigma)/(2sigma),sigma="+svm.svm_get_svr_probability(model)+"\n");
        }
        else
        {
                svm.svm_get_labels(model,labels);
                prob_estimates = new double[nr_class];
                output.writeBytes("labels");
                for(int j=0;j<nr_class;j++)
                        output.writeBytes(" "+labels[j]);
                        output.writeBytes("\n");
        }
}
while(true)
{
        String line = input.readLine();
        if(line == null) break;
        StringTokenizer st = new StringTokenizer(line," \t\n\r\f:");
        double target = atof(st.nextToken());
        int m = st.countTokens()/2;
        svm_node[] x = new svm_node[m];
        for(int j=0;j<m;j++)
        {
                x[j] = new svm_node();
                x[j].index = atoi(st.nextToken());
                x[j].value = atof(st.nextToken());
        }
```

Converting input format in order SVM module to read

Reading word vectors

Fig. 10. Programming Code for Converting Input Format and Reading Word Vector

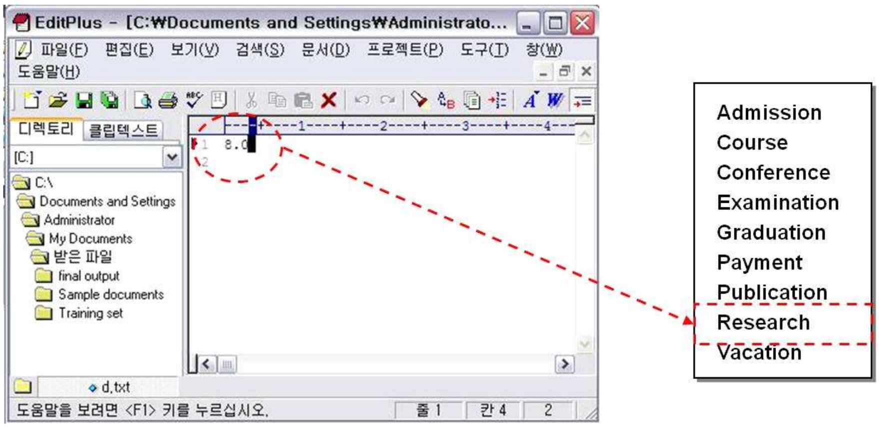

Fig. 11. The result of SVM-based classification

the rating points in Table 1 show. Based on the results in the Table 1, an interesting but very important implication can be derived: Key requirements to make the knowledge acquisition process be fully-automated can be identified. 'STT App-based dictation' and

```
// move input_document(dictated conversation) to resultant category====================

        // read the result category no
        BufferedReader result = new BufferedReader(new FileReader("temp_for_libsvm/KM_OutputFile"));
        String resultCategoryNo = result.readLine();
        String resultCategoryTitle = null ;

        // find destination directory
        for ( int at = 0 ; at < kmCategoryInfo.size() && resultCategoryNo != null ; at++ ){
                categoryInfo tempcate = (categoryInfo)kmCategoryInfo.elementAt(at);
                if ( resultCategoryNo.equals(tempcate.getCategoryNo() ) ){
                        resultCategoryTitle = tempcate.getCategoryTitle() ;
                        break ;
                }
        }

        // copy file to destination category
        int tempi, len=0;
        InputStream in = null;
        OutputStream out=null;

        try{
                in= new FileInputStream(new File("input_document/"+argv[0]));
                out= new FileOutputStream("classified_documents/"+resultCategoryTitle+"/"+argv[0],false);
                while((tempi=in.read())!= - 1){
                out.write(tempi);
                len++;
        }

 in.close();
 out.close();

        //  delete input file
        File delFile = new File ("input_document/"+argv[0]);
        System.out.println(delFile.getAbsoluteFile()  );
        if( !delFile.delete() ) System.out.println("file is not deleted");
        }catch(Exception ee){System.out.println(ee.getMessage() );}
```

Fig. 12. Programming Codes to Store Dictated Conversation('input_document') under the Identified Topic Performance

'SVM-based topic identification' are the key features enabling the prototype system to be regarded as a fully-automated system.

The feature of 'STT App-based dictation' can give an answer to the question of how knowledge in ordinary conversations and meetings can be acquired. Knowledge in conversations is too various to count the number, and indeed is street smart, which makes researchers and practitioners have long concentrated on it. Voice knowledge has been regarded as something that could not be articulated and acquired, however the functionality of STT App-based dictation provides a way convenient to manage voice knowledge. Also, the feature of 'SVM-based topic identification' makes the prototype system possible to acquire knowledge with understanding the subject included in the knowledge. Dictating knowledge without knowing its meaning cannot be regarded as 'acquisition of knowledge' [4]. Dictating voice knowledge and identifying its topic must be paired together to attain the true acquisition of voice knowledge, which is one of primary contributions of this research.

4.3.2 Statistical Analysis to Measure the Accuracy of Acquisition

The accuracy of acquisition can be influenced by two components: STT application and SVM module. The accuracy of STT App-based dictation is very crucial, because it initiates the execution of knowledge acquisition process. Unless the accuracy of dictation is guaranteed, the result of acquisition cannot be satisfactory. The accuracy of STT App-based dictation, however, depends on the speaking habit of speakers, not on the recognition ability of applications. Of course the function of recognition must be enhanced so that the application can recognize every pattern of speaking; however this work must be done by researchers in the area of voice recognition technologies. Considering the issue of recognition accuracy is out of this research's concern. Although the accuracy of STT App-based dictation was observed to be very high, even a little error could have bad effects on the final results. Therefore, in this research, assuming that the performance of STT application is satisfactory enough not to have any effects on the final results is more reasonable.

On the contrary, the accuracy of the SVM module can and must be considered in this research. The accuracy of the SVMs has been verified to be very high. If the prediction model has been trained sufficiently, then the SVMs output very accurate and correct results. The accuracy can refer two aspects: one is for comparing manual classification and the other is for measuring the correctness of classification. The LibSVM deployed in this research has been proved to outperform the manual classification [8]. Measuring the correctness of classification means how much the outputs are classified into the correct categories, and usually it can be measured in statistical considerations. To measure the correctness of classification, three to five documents per each category are collected, and totally 40 test documents are inputted to check the accuracy. Figure 13 shows the prediction accuracy which is automatically calculated by the LibSVM. The accuracy of classification is indicated to be 82.5% with 1.025 MSE. This means 33 documents out of 40 documents are correctly classified

Table 2. Performance Evaluation by Focus Group Interview

Member	# of Trials	Rating Pts	Features		
			Fully-automated	**Semi-automated**	**None-automated**
1	3	6	- STT App-based dictation - Topic identification - Knowledge storing	- Data transmission	
2	3	6	- STT App-based dictation - Topic identification	- Knowledge storing	
3	5	7	- STT App-based dictation - Data transmission - Topic identification - Knowledge storing		
4	5	6	- STT App-based dictation - Data transmission - Topic identification	- Knowledge storing	
5	2	5	- STT App-based dictation - Data transmission - Topic identification	- Classifier training	- Business rule extraction
6	3	6	- STT App-based dictation - Topic identification - Knowledge storing		
7	5	6	- STT App-based dictation - Topic identification	- Knowledge storing	
8	3	6	- STT App-based dictation - Topic identification		
9	5	5	- STT App-based dictation	- Topic identification - Knowledge storing - Data transmission	- Business rule extraction - User confirmation
10	4	7	- STT App-based dictation - Topic identification - Data transmission		

into proper categories. Comparing to the results of the previous studies [8, 10], this result shows that the SVM module of this research classifies accurately. Of course, the accuracy needs to be improved around 90% to gain the credibility of classification. Small number of sample documents for training is the main reason for the relatively low level of accuracy. If sufficient number of training documents is provided in training the model, more accurate result of classification can be expected.

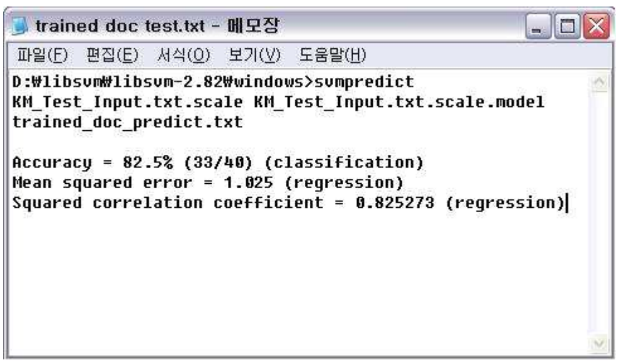

Fig. 13. Prediction Accuracy of the SVM module

5 CONCLUSIONS AND FUTURE WORKS

The question "*Do you know what you do and do not know?*" has long been issued by the researchers of knowledge management. Although one may answer affirmatively to the question, he or she would have considerable difficulty in answering the following question: "*Can you articulate all of them?*" The reason why knowledge cannot be effectively accumulated is not because we do not know what we know, but because we do not know how we articulate it. If there is way to document we know, then a tremendous amount of knowledge can be identified and accumulated in knowledge repositories. The limitations in the methodology of knowledge acquisition have restricted the practices of knowledge management. After something to be managed have been prepared, specific technologies to manage them can finally be valuable.

This paper tries to provide an effective solution to this problem by truly automating the process of knowledge acquisition. The amended knowledge acquisition methodology, proposed by this paper, enhances the conventional approaches to automatically acquire knowledge by applying the concept of autonomous computing and the Smartphone operated on the cloud computing environment. The context of knowledge use is also considered using the Smartphone as a sensor to identify context data. To promote the use of acquired knowledge, the context data which play the role of the meta-knowledge must be considered.

Among various types of knowledge, this research just focuses on the voice knowledge which verbally communicated in conversations. Knowledge in conversations is to be acquired by identifying the topic of an electronic document containing the dictated texts which have been converted from the recorded conversations. Therefore, other types of knowledge, such as hard copy- and activity-based knowledge, can be also acquired by applying the same procedure if, and only if, the electronic document is provided. The optical character recognition technology can be applied to convert the hard copy document into the electronic document, and the sensor-based activity recognition technology can be applied to convert the monitored actions into the textual data. Since knowledge can make its appearance to the real world via conversations, documents, and activities, observing such human outputs and applying the proposed methodology to them can deliver a full-scale management of knowledge.

The SMART-ConKAS, a prototype system based on the proposed methodology, includes two main sub-modules that are synchronized: the STT module and the SVM module. Therefore, the accuracy of the SMART-ConKAS's output (i.e., its performance) depends solely on the performance of each sub-module. If the STT module cannot correctly recognize the knowledge holder's speech, then outputs are naturally based on wrong or unidentifiable topics. If the SVM module is not trained sufficiently, then wrong outputs are naturally extracted. Thus, the quality of the sub-module determines the performance of the SMART-ConKAS. Therefore, the determination of the most appropriate sub-module for the SMART-ConKAS is a critical issue in this paper. Of course, ensuring correct outputs from each sub-module can address this problem, but this paper leaves this topic to future research because it is beyond the scope of this paper.

A discussion can include several pieces of knowledge. Therefore, a sentence-by-sentence analysis is necessary to clearly differentiate one piece of knowledge from another. However, the SMART-ConKAS is designed to simply store the whole discussion in terms of extracted topics. Although a discussion can be considered as a piece of knowledge [1], each piece must be distinctively identified for its realistic use. For this, sentences within a discussion must be segmented, and each sentence must be analyzed in terms of its accidence. In examining each part of a sentence, knowledge-resident parts must be analyzed by expert linguists. Although this paper also assumes that a discussion is a piece of knowledge, there should be a linguistic analysis of every sentence to ensure the reliability of acquired knowledge.

In addition, the absence of a universal as well as formal corpus of contexts must be resolved. Such a corpus (namely predefined context categories) can play an important role in labeling each branch of the ontology hierarchy, and thus, the number of words that a system can understand varies according to the corpus. This paper simply sets a corpus based on a project-reported category because there is no formal and universal corpus appropriate for the paper. Although there is Reuter-21578, a widely used corpus, it cannot be applied to this paper because of differences in the way in which the subjects are described. A number of researchers focusing on ontology development have attempted to construct a formal and universal corpus, and thus, the SMART-ConKAS should include a more formal and universal corpus for the more accurate and realistic classification of knowledge.

This paper contributes to the knowledge management literature by suggesting a new methodology to automatically acquire human knowledge based on the capabilities of autonomous and cloud computing technologies. This study is different from conventional studies in that it provides and implements a fully-automated knowledge acquisition methodology. Organizations seeking to manage organizational knowledge for strategic purposes can apply the proposed approach to their business environment. As long as knowledge possessors are agreeable to sharing and monitoring their knowledge, the proposed methodology and the prototype system should facilitate organizations' knowledge acquisition efforts.

In particular, this study extends ubiquitous computing to the knowledge acquisition process. The results suggest that ubiquitous computing can not only facilitate a more user-friendly computing environment but also enable a more human-centered society. In addition, by applying the concept of context to the knowledge management, this study identifies a way to enhance the quality of knowledge and makes it more context-oriented. Based on context-rich knowledge, ordinary activities must become more convenient and economic through the minimization of mistakes and costs.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation (NRF) of Korea funded by the Ministry of Education, Science and Technology (No.H00021).

REFERENCES

[1] Anerousis, N. and Panagos, E., "Making Voice Knowledge Pervasive", *IEEE Pervasive Computing*, Vol.1 No.2, pp.42-48, 2002.

[2] Basu, A., Watters, C. and Shepherd, M., "Support Vector Machines for Text Categorization", *Proceedings of the 36th Hawaii International Conference on System Sciences*, 2002.

[3] Chang, C.C. and Lin, C.J., *LIBSVM: a library for support vector machines*, available at http://www.csie.ntu.edu.tw/˜cjlin/libsvm, 2001.

[4] Davenport, T.H. and Prusak, L., *Working Knowledge: How organizations manage what thet know*, Harvard Business School Press, 1998.

[5] Dumais, S., Platt, J., Heckman, D., and Sahami, M., "Inductive learning algorithms and representations for text categorization", *Proceedings of the 7th International Conference on Information and Knowledge Management*, 1998.

[6] Edwards, J.S., Alifantis, T., Hurrion, R.D., Ladbrook, J., Robinson, S., and Waller, A., "Using a simulation model for knowledge elicitation and knowledge management", *Simulation Modelling Practice and Theory*, Vol.12, pp.527-540, 2004.

[7] Huang, W. and Tao T., “Adding Context-awareness to Knowledge Management in Modern Enterprises”, *Proceeding of the 2nd IEEE International Conference on Intelligent Systems*, pp.393-398, 2004.

[8] Hsu, C.W., Chang, C.C., and Lin, C.J., *A Practical Guide to Support Vector Classification: LibSVM Tutorial*, available at http://www.csie.ntu.edu.tw/˜cjlin/papers/guide/guide.pdf, 2001.

[9] Joachims, T., “Text categorization with support vector machines: Learning with many relevant features”, *Proceedings of the European Conference on Machine Learning*, 1998.

[10] Kang, H., Suh, E., and Yoo, K., “Packet-based context aware system to determine information system user’s context”, *Expert Systems with Applications*, Vol.35 No.1-2, pp.286-300, 2008.

[11] Kwan, M.M. and Balasubramanian, P., “KnowledgeScope: managing knowledge in context”, *Decision Support Systems*, Vol.35, pp.467-486, 2003.

[12] Kwon, O. and Lee, J., “Text categorization based on k-nn approach for web site classification”, *Information processing and management*, Vol.39, pp.25-44, 2003.

[13] Meyer, D., Leisch, F. and Hornik, K., “The support vector machine under test”, *Neurocomputing*, Vol.55, pp.169-186, 2003.

[14] Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., and Euler, T., “YALE: Rapid Prototyping for Complex Data Mining Tasks”, *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-06)*, 2006.

[15] Nunes, V.T., Santoro, F.M., and Borges, M.R.S., “A context-based model for Knowledge Management embodied in work processes”, *Information Sciences*, Vol.179, pp.2538-2554, 2009.

[16] Pallis, G., “Cloud computing: the new frontier of Internet Computing”, *IEEE Internet Computing*, Vol.14 No.5, pp.70-73, 2010.

[17] Rennie, J.D.M. and Rifkin, R., “Improving multiclass text classification with the support vector machine”, CBCL Paper #210/AI Memo #2001-026, Massachusetts Institute of Technology, Cambridge MA, October, 2001.

[18] Schmidt, A., Beigl, M., and Gellersen, H., “There is more to context than location”, *Computers and Graphics*, Vol.23, pp.893-901, 1998, In Zhou, J., Gilman, E., Palola, J., Riekki, J., Ylianttila, M., and Sun, J., “Context-aware pervasive service composition and its implementation”, *Personal and Ubiquitous Computing*, Vol.15, pp.291-303, 2011.

[19] Takashiro, T. and Takeda, H., “A Context Based Approach to Acquisition and Utilization of Personal Knowledge for WWW Browsing”, *Proceedings of the 4th International Conference on knowledge-Based Intelligent Engineering Systems & Allied Technologies*, pp.756-759, 2000.

[20] Voida, S., Mynatt, E.D., MacIntyre, B., and Corso, G.M., “Integrating Virtual and Physical Context to Support Knowledge Workers”, *IEEE Pervasive Computing*, Vol.1 No.3, pp.73-79, 2002.

[21] Yan, H., Jiang, Y., Zheng, J., Fu, B., Xiao, S., and Peng, C., “The internet-based knowledge acquisition and management method to construct large-scale distributed medical expert systems”, *Computer Methods and Programs in Biomedicine*, Vol.74, pp.1-10, 2004.

[22] Yang, K. and Huh, S., “Automatic expert identification using a text categorization technique in knowledge management systems”, *Expert Systems with Applications*, Vol.34, pp.1445-1455, 2008.

[23] Yoo, K., “SVM-based knowledge topic identification toward the autonomous knowledge acquisition”,*Proceedings of the IEEE 9th International Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pp.149-154, 2011.

[24] Yoo, K., “Autonomous and pervasive computing-based knowledge service”, *Lecture Notes in Electrical Engineering*, Vol.156, pp.9-14, 2013

[25] Zhou, J., Gilman, E., Palola, J., Riekki, J., Ylianttila, M., and Sun, J., “Context-aware pervasive service composition and its implementation”, *Personal and Ubiquitous Computing*, Vol.15, pp.291-303, 2011.

Keedong Yoo is an associate professor in the Department of MIS at Dankook University, South Korea (kdyoo@dankook.ac.kr). He has B.S. and M.S. in Industrial Engineering from the POSTECH (Pohang University of Science and Technology), South Korea; and a Ph.D. in Management and Industrial Engineering from the POSTECH. His research interests include knowledge management and service; intelligent and autonomous systems; context-aware and pervasive computing-based knowledge systems.

AUTHORS’ ADDRESSES

Asst. Prof. Keedong Yoo, Ph.D.
Rm238 Bldg of Social Science,
Dankook Univeristy(Cheonan Cam),
119 Dandae-ro, Dongnam-gu, Cheonan-si, Chungnam,
Republic of Korea (330-714),
email: kdyoo@dankook.ac.kr

Received: 2012-11-09
Accepted: 2013-01-08