Holistic Framework and Application Suite for the Knowledge Service: Focused on the Management of Verbal Knowledge

*Keedong Yoo
Dept. of MIS, Dankook University, kdyoo@dankook.ac.kr

Abstract

Knowledge must be acquired and distributed in a consistent and integrated manner without user's direct interference to save and use it. Most of conventional approaches of the knowledge service, however, have fragmentally considered either of them ignoring user's integrated way of knowledge use. This paper introduces a holistic framework of the knowledge service simultaneously encompassing both of automated knowledge identification and distribution, especially focused on the knowledge externalized via dialogues. A smartphone-based prototype automatically capturing, storing, and distributing verbal knowledge is also designed and implemented to validate the proposed framework. This paper's ideas can contribute enhancing functionalities of intelligent personal assistant applications recently gaining users' much interest to be more user-oriented and effective.

Keywords: *Knowledge service, Automated knowledge identification, Automated knowledge distribution, Verbal knowledge, Application suite.*

1. Introduction

Users, nowadays, tend to solve or fix daily problems by themselves as the IT-based service for generating and providing knowledge are developed and proliferated, and therefore the service of knowledge is becoming one of basic requirements for the modern life. The service of knowledge, therefore, must be delivered to users accurately and automatically to meet their needs in quality and quantity of knowledge. Using recent mobile devices and the Internet-related technologies, current users become more knowledgeable than ever, which upgrades users as outlanders to users as experts [1, 2]. Recently released knowledge services and related devices, such as Apple's Siri (apple.com/ios/siri/), Google's Now (androidcentral.com/google-now), Microsoft's Cortana (microsoft.com/en-us/mobile/experiences/cortana/), and Amazon's Echo, etc., are typical examples demonstrating such trends. Using these devices and applications, users can enjoy voice-based commanding, simple dialoging, and rule-based automatic device activating services.

A smartphone is the very effective device experiencing the knowledge service not only in size and weight, but also in instantly writing and extracting knowledge [3]. The smartphone users can store what they know, and can monitor what they need to know at any time and places using smartphone's functions of recording and searching. Using the capabilities a smartphone has, in other words, knowledge can be effectively captured and distributed, and therefore users can gain more experiences from the knowledge service. However, no ongoing knowledge services utilizing smartphone's potential to automatically capture knowledge can be found. By applying smartphone's functions of writing, recording, and photo-taking, the service of knowledge identification can be added to current way of knowledge services. The knowledge life cycle, captured, stored, shared, utilized, and upgraded knowledge, can be fully managed, and therefore users' practical desire toward managing knowledge by themselves can be achieved by the knowledge service.

This paper, therefore, proposes a framework of the knowledge service holistically encompassing capabilities of knowledge identification and distribution using a smartphone. Among various types of

---

* Corresponding Author
Received: Mar. 16, 2017, Revised: Mar. 25, 2017, Accepted: Mar. 27, 2017

knowledge, verbally expressed knowledge during dialogues, namely the verbal knowledge, is selected to be exemplified, because smartphone's recording and dictating functions are proper examples to show its potentials to the knowledge service. To prove the proposed concept's validity, a prototype can be regarded as an application suite for the knowledge service is illustrated by screen shots and sample programming codes.

2. Knowledge Service Framework

The proposed framework automatically identifying and distributing knowledge consists in four kinds of sub systems, topic identifier, knowledge archiver, cloud storage-based knowledge base, and knowledge extractor, as Figure 1 shows. The contents of a dialogue recorded and dictated from the user's smartphone is transmitted to the topic identifier which determines the topic of inputted file. Once the topic is concluded, it is tagged on the dictated dialogue with other meta-knowledge, such as the ID of dialogue participants, the time of recording, and the location of meeting, etc., to form a set of knowledge. Resultant knowledge set is transmitted to the knowledge archiver determining target storing directory. After determining proper directory to store inputted knowledge set, the knowledge archiver transmits the knowledge with directory information to the knowledge base. Knowledge identification process is completed by archiving inputted knowledge under the knowledge base which changes the status of archived one to be stand-by for being requested by knowledge distribution process. Knowledge distribution is initiated by the call of the knowledge extractor which triggers proper knowledge based on analyzed, namely inferred, user contexts or their direct requests. By sending and displaying extracted knowledge onto user's smartphone, the whole processes of the knowledge service is finished: A holistic procedure for knowledge identification and distribution is completed. Specific explanation about each subsystem is as follows;

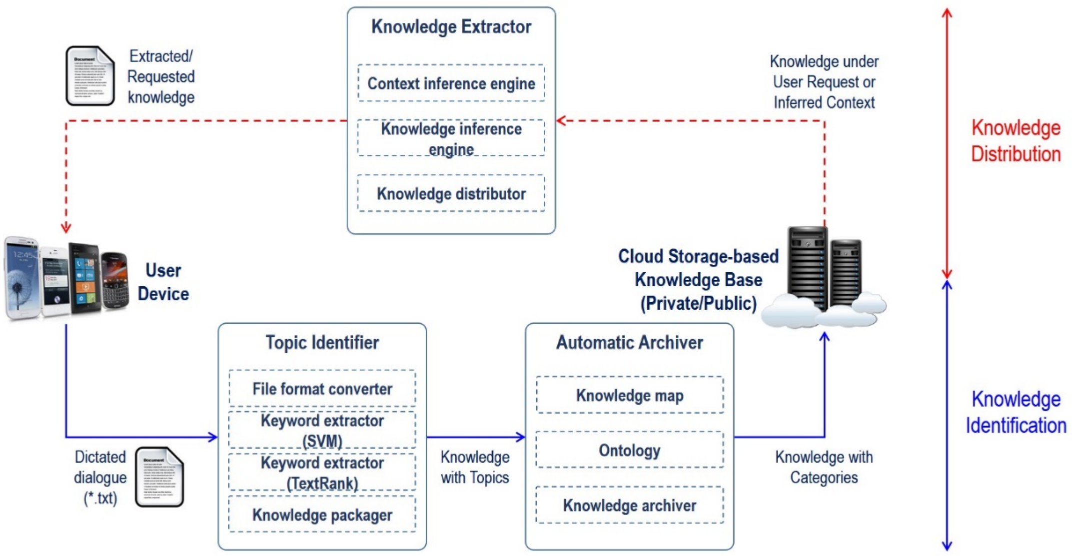


Figure 1. Proposed framework for the knowledge service.

2.1 Subsystem #1 for Knowledge Identification: Topic Identifier

The topic identifier determines the topic of dictated dialogues transmitted from user's smartphone by extracting their keywords or keyphrases. To do this job, the file format of dictated dialogues must be converted first so that a keyword extractor can read them. Following programming codes show how to convert file formats, especially for the case that the SVM(Support Vector Machine) is applied as a keyword extractor.

```
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());
      }
```

Then the keyword extractor analyzes inputted file to determine highly ranked keywords or keyphrases. In this paper, two approaches to do this job are suggested: SVM as a supervised method and TextRank as an unsupervised one. SVM is known extracting keywords very accurately, and therefore widely used in the area of supervised learning-based text mining [4]. However, as do most of supervised methods, since SVM requires enough size of pre-training data to accurately predict the vector space of inputted document. Worse than this, because the language of pre-training data must coincide with that of inputted document, the language type-dependent corpus, the categorized collection of document, must be prepared in advance. To meet with this problem, the TextRank, one of widely used unsupervised methods, can be alternatively applied. The TextRank exploits the structure of the word itself to determine keyphrases that appear 'central' to the word in the same way that PageRank selects important Web pages. Because the TextRank enables the application of graph-based ranking algorithms to natural language texts, it produces results independent to the training data and language types [5, 6]. Considering the efficiency of keyword extraction, the TextRank is more reasonable; considering the effectiveness, The SVM is recommended. In this study, two approach are alternatively applied so that the TextRank performs its job first to form enough trial sets possibly used as training data for activating the SVM.

Then extracted keywords or keyphrases are sent to the knowledge packager tagging extracted results on the inputted file as the topic. To avoid the case that different files have the same topic, additional information, such as the ID of dialogue participant, the time of recording, and the location of meeting, are also tagged on dictated dialogue as meta-knowledge. Therefore, one set of knowledge can be finally composed.

2.2 Subsystem #2 for Knowledge Identification: Knowledge Archiver

The knowledge archiver stores transmitted set of knowledge and meta-knowledge into the knowledge base according to previously identified keywords or keyphrases. For the extended use of knowledge, newly emerged knowledge must be stored considering the relationships with existing knowledge, and the knowledge map plays the key role to do this job. Since the keywords of knowledge to be stored are automatically determined, the link between related knowledge that has the same keywords can be also automatically made. By expanding links between knowledge, a knowledge network can be formulated, and this is the knowledge map. Using a knowledge map, users can utilize knowledge associatively in the way of following links in the knowledge network. Because the directories of the knowledge base are defined according to the structure of a knowledge map, they have the form of a network, which is complex but very effective in self-reasoning, if network-typed knowledge map is prepared first. To define and identify relationships between knowledge, the ontology technology is strongly recommended. Using the ontology technology, hierarchical directories can be easily transformed to network-typed ones, and, in addition to this, the directories can be expanded automatically by ontology's self-reasoning.

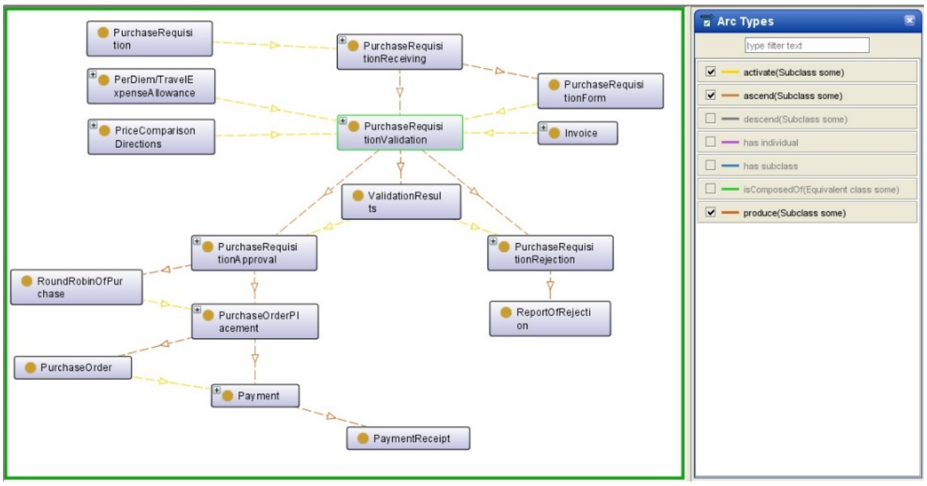


Figure 2. Example of ontology-based knowledge map.

Figure 2 shows the example of an ontology-based knowledge map. Once the directory for identified knowledge is determined, the knowledge archiver transmits knowledge and its meta-knowledge data together into a knowledge base. Following codes show how knowledge and meta-knowledge can be combined and transmitted together.

```
public KBclientsocket (Knowledge theKnowledge, KBContext theContext, String
KBhostname) throws IOException {

	// connnecting a KB
	Socket KBSocket = new Socket(KBhostname, KBServerPort);
	// transmitting knowledge and its context data
	ExchangeKB(theKnowledge, KBContext, KBSocket) ;
	// disconnecting
	KBSocket.close();
}
public class KBContext{
	string ID;
	string Time;
	string userLocation;
}
```

2.3 Subsystem #3: Cloud Storage-based Knowledge Base

Using benefits of the cloud storage, the knowledge base stores transmitted knowledge at any time and place. Since the knowledge base must also respond to the knowledge request from users, it must equip the cloud storage-based ubiquitous capability. The knowledge base can be composed of two kinds of split repositories: the public and private knowledge base. While the public knowledge base is usually used in workplaces officially legislated and centrally controlled, the private knowledge base usually manages personal, and so very private knowledge assets. Conventionally emphasis was given to the public knowledge base only, however, importance of the private knowledge base is being increased because of users' desire to manage knowledge by themselves.

2.4 Subsystem #4 for Knowledge Distribution: Knowledge Extractor

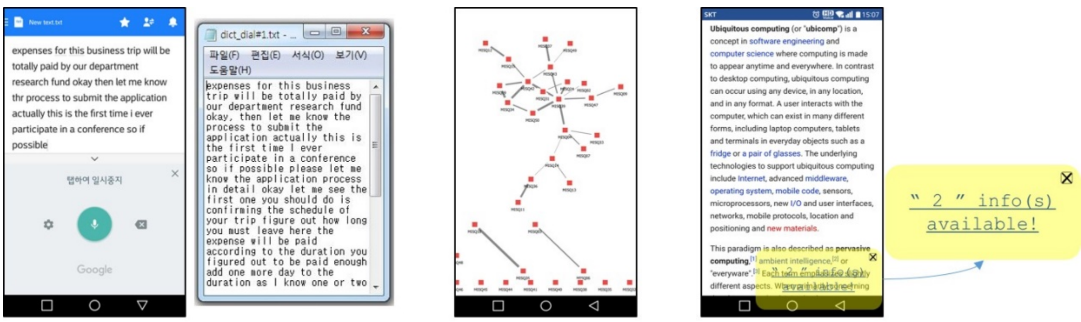
The knowledge extractor selects, extracts, and transmits knowledge proper to users' direct requests or to user's inferred situation. Using a smartphone, user's context data, explains current user's environment, can be efficiently captured, and therefore it is possible to automatically provide knowledge timely relevant to user's situation. User's identity, time, place, and temperature are typical example context data used in determining user's situation. The contents displayed in a PC monitor or a smartphone are also able to be applied to conclude user's mental situation by thinking [7, 8]. Following codes show how to extract and transmit knowledge according to concluded user's situation (context).

```
public Knowledge getKnowledgeByContext (KBContext theContext, String KBhostname)
throws IOException {

      Knowledge result;
      // socket connection to KB
      Socket KBSocket = new Socket(KBhostname, KBServerPort);
      // sending knowledge proper to context
      if(checkKB(theContext, KBSocket)) result = receive(theContext, KBSocket) ;
      else result.setKnowlegeText("not found");
      KBSocket.close();
      return result;
}
```

3. Example Prototype of Application Suite for the Knowledge Service

Based on the proposed framework, prototype applications are can be implemented. An application suite for the knowledge service can be constructed, if every module explained in proposed framework is implemented and integrated, because it deals both of knowledge identification and distribution. As explained previously, this study deploys a smartphone as the main user device to identify and request knowledge, and therefore verbal knowledge dictation and knowledge request are designed to be performed using a smartphone. Sample screen shots of selected functions that can be externally exhibited are shown in Figure 3.



(a) (b) (c)

Figure 3. Screen shots of implement prototype
((a) Recording and Resultant dictated dialogue, (b) Ontology-based knowledge map, (c) Knowledge(information) recommendation)

4. Concluding Remarks

This paper suggests a holistic framework for the knowledge service by integrating the processes of knowledge identification and distribution. Comparing conventional approaches mainly paying attention to the process of knowledge distribution, this study emphasizes the importance of automating the process of knowledge identification also to meet with current users' needs on the knowledge service. The performance of the implemented prototype is not satisfactory yet, because of several technical issues such as the inaccuracy of dictation, not enough and proper training data of the keyword extraction algorithm, and irrelevant recommendation of knowledge, etc. Those limitations must be resolved as each area of voice recognition, machine learning, and context inference progresses. Smartphone's limited usability because of its small display size must be also reinforced to effectively attract users' interests [9, 10]. In spite of such practical limitations, this paper's value can be found in including the process of knowledge identification into the knowledge service framework, and in suggesting logical relationships between component subsystems. This paper's ideas can underpin preparing a standard platform of the knowledge service, and therefore promote more knowledge-based lives of users.

5. Acknowledgements

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2013S1A5A2A01017530).

6. References

[1] Berry, P.M., Gervasio, M., Peintner, B., and Yorke-Smith, N. (2011). PTIME: Personalized assistance for calendaring. ACM Transactions on Intelligent Systems and Technology. 2(4), 40.

[2] Yoo, K. (2013). Capture knowledge on the spot: toward the autonomous and pervasive Service of context-rich knowledge. Automatika. 54(4), 401-414.

[3] Chon, J. and Cha, H. (2011). LifeMap: A Smartphone-Based Context Provider for Location-Based Services. IEEE Pervasive Computing. 10(2), 58-67.

[4] Meyer, D., Leisch, F. and Hornik, K. (2003). The support vector machine under test. Neurocomputing. 55, 169-186.

[5] Mihalcea, R. and Tarau, P. (2004), "TextRank: Bringing order into texts", in Proceedings of EMNLP, 4(4).

[6] Cui, W., Wang, P., Du, Y., Chen, X., Guo, D., Li, J., and Zhou, Y. (2017). An algorithm for event detection based on social media data. Neurocomputing. In press.

[7] Kim, S., Suh, E., and Yoo, K. (2007). A study of context inference on the Web-based information systems, Electronic Commerce Research and Applications. 6, 146-158.

[8] Van Doorn, E., Rusák, Z., and Horváth, I. (2017). A situation awareness analysis scheme to identify deficiencies of complex man-machine interactions. International Journal of Information Technology and Management. 16(1).

[9] Ahmad, N., Boota, M.W., and Masoom, A.H. (2014). Smart Phone Application Evaluation with Usability Testing Approach. Journal of Software Engineering and Applications. 7, 1045-1054.

[10] Jung, N. and Jang, M. (2016). Effects of screen size on smartphone functionality and usability for stroke patients with hemiparalysis. Journal of Physical Therapy Science. 28(4), 1330–1334.