A Context-based Support System of Mobile Chinese Learning for Foreigners in China

Lijun Sun\textsuperscript{a,*}, Jing Hou\textsuperscript{a}, Xiangpei Hu\textsuperscript{a}, Khalil Al-mekhlafi\textsuperscript{a}

\textsuperscript{a}Institute of Systems Engineering, School of Management, Dalian University of Technology, Dalian 116023, Liaoning, China

Abstract

Taking advantage of the improving Mobile Web technologies, mobile language-learning systems plays an important role for self-learning, especially in the case of working people. Nevertheless, existing systems do not support learners in studying or practicing language materials related with the current context. As it is known, the context related with surroundings can deepen the impressions and naturally explain the meanings of the sentences in a learner’s memory, which leads to an easy and casual language-learning experience. To achieve this goal, a context-based support system of mobile Chinese learning for foreigners in China is proposed by this research. The system can provide a learner with Chinese-learning materials related with what he is doing or what he is going to do. Besides, it also can generally adapt to learners’ profiles like many existing personalized e-learning systems. In the system, a context-based ontology is investigated to represent the knowledge that is used to support the process of matching context with suitable language-learning materials, and inference rules in the matching process are developed by Semantic Web Rule Language (SWRL). Four steps of the matching process are realized based on the knowledge representation method and the rules. The implementation of the prototype system shows that it can support the mobile Chinese-learning process by providing a learner with suitable learning materials related with the contexts of the task he/she is doing or going to do, his/her profile and mobile device, so that it can greatly improve the effect and efficiency of learning.

1. Introduction

With the development of economy in China, more and more foreigners come to China for trading, living or...
studying, so the demand for learning Chinese is surging. For foreigners who are busy with other businesses all day long and cannot allocate special time for studying Chinese, accessing learning materials at any time and at any place becomes important so that they can study or practice at their spare time. Though electronic textbooks have been popular, which can be stored in portable devices and accessed at any time and at any place, this kind of e-dictionary-like language learning cannot relate the content of learning materials to the real-world context of a learner. However, relating learning materials to the context is very important, because learners can achieve learning effect by communicating with Chinese natives while they finish some daily tasks of their own, which is in accordance with the concept of situated learning and task-based learning in second language acquisition. The theory of situated learning holds that learning itself is the outcome of mutual interactions between learners and his environment, not just about acquiring knowledge. The theory of task-based learning means that learners will be assigned several specially designated tasks correlated to certain learning objective. While in the process of finishing these tasks, the learner’s capabilities will also grow up.

Therefore, in this paper, we propose a context-based support system of mobile Chinese learning, which can support a learner in studying or practicing the learning content related with the current context. In other words, the system can provide a learner with the Chinese-learning materials related with what he is doing or what he is going to do. It’s a context-based language-learning system. The context related with surroundings can deepen the impressions of the sentences in a learner’s memory. Moreover, under the communication in real-world environment, the deeper cultural implications in Chinese language can be experienced in daily life. Consequently, the effect and efficiency of the learning will be greatly improved. Another advantage of the system is that it is implemented in a mobile device and suitable for being used by a mobile phone. Mobile phones are popular now. Almost everyone owns one, which means that the system is available almost for everyone so that it will make the learning of Chinese greatly popular.

To develop this kind of system, the core of the research should be how suitable learning materials are provided to a learner based on his/her current context. As context changes in real time with the mobility of a learner and different contexts may relate with different contents of learning materials, the keyword-based matching method cannot satisfy the dynamic matching between contexts and learning materials. A new matching process based on context should be investigated for implementing the system.

The remaining part of the paper is organized as follows. In section 2, the work related with mobile language-learning systems is reviewed. In section 3, the architecture of the system is presented firstly. In order to realize the functions of the system, the context related with language learning is analyzed. Based on the analysis, the knowledge representation method for supporting the matching process between context and language-learning materials is investigated, and finally, the matching process is studied. In section 4, a scenario that the learning process is related with the task of completing a deposit in a bank demonstrates how the system is implemented and used. Conclusions are drawn finally with the further study of the research in section 5.

2. Related work

Existing mobile language-learning systems can be categorized into four kinds, the system of electronic dictionary, the system of automatically sending language-learning short messages, the system of linking students with language-learning classes or with teachers, and the system of a specific language-learning topic.

(1) The system of electronic dictionary is the extension of traditional e-dictionaries. This kind of system just has the functions of storage and search based on keywords or navigation labels. For example, Duke University adopts iPod as the assistant of teaching. The functions include course content dissemination, classroom recording, field recording, study support, and file storage and transfer. Tseng, Lu, and Hsu proposed a system that works on PDA for foreigners to learn Chinese. The system imagines that the electronic vocabularies or dialogues can be retrieved according to the GIS information, which attempted to combine the geographical context into Chinese learning. In the commercial area, the NTT DoCoMo Company in Japan provides English-Japanese dictionary, Japanese-English dictionary, and grammars and hot vocabularies with low cost by its Pocket Eijiro website. The amount of persons subscribing these electronic language-learning materials is over ten thousand. Nokia also launched a software that supports mobile English learning. It can be applied by 15 types of Nokia mobile phones. The learning materials are
(2) The system of automatically sending language-learning short messages is the extension of the traditional class-based language-learning method. Teachers will prepare short messages with learning content and set a specific time to send them to students. For example, Mike and Claire extended the use of Telstra Mobile Online SMS Business service to mobile Italian language learning, which initially just allows SMS text messages to be prepared in advance and then sent to a large group of people via e-mail and PC. After extension, a message may be prepared in advance and then scheduled to be sent to a group of students’ mobile phones on a certain day at a certain time in the future. Messages may also be sent ‘Once Only’ or as a ‘Recurring Delivery’, in which case the sender can nominate the intervals at which the message will be sent again. Thornton and Houser did an experiment in Japan to test the learning effect by SMS for students of English as a Foreign Language (EFL). They found that delivery of foreign language vocabulary lessons via mobile phone e-mail is effective and received positively by Japanese university students. This method takes advantage of the push aspect of mobile technology, and promotes regular study.

(3) The system of linking students with language-learning classes or with teachers sets the communicating environment between a learner and a class or a teacher. This kind of system cannot only provide the electronic materials, but also provide an environment for students to communicate with language teachers or with classes in real time. The researches on this kind of system include the teaching model, the method of material collection and dissemination. Luo constructed an English-learning platform with this kind of subsystem. Wang developed a mobile English-learning system that provides a communicating mechanism between students and teachers, by which the teacher can guide students in distance. In L&S Learning Support Services at the University of Wisconsin-Madison, HP Jornada mobile devices were tested in the aspect of supporting mobile language learning by three experiments of Norwegian, French & Italian, and Latin Literature. This kind of devices allowed students to engage in open-ended communication about more complex ideas than they could handle in oral discussion.

(4) The system of a specific language-learning topic just provides learning materials or provides the support for mobile learning activities on a specific language topic. For example, Tao investigated the development of a mobile English-learning software on the topic of Olympic English. This system provides functions of learning vocabularies, phrases and sentences, reading passages, practicing dialogues, online discussing, and online updating of Olympic English. Zhu developed a mobile learning system on the topic of English for the title of a technical or professional post by using the technology of Java ME. Learning materials on this topic can be scanned online or downloaded through the wireless network and stored in a mobile phone for learning at anytime. This system was tested and evaluated in the two types of mobile phones, Nokia A3230 and Dopod D9000.

These mobile language-learning systems mentioned above improved the technologies for mobile language learning, but most of them cannot utilize the situated factors to assist the learning process. Combining context or context-aware technologies into mobile language-learning systems to utilize situated factors in the process of language learning should be further studied.

3. Context-based support system of mobile Chinese learning

The users of the proposed system are foreigners studying, living or working in China, who’s native or second language is English. We assume that their English levels are good enough for daily use. The proposed system adopts English not Chinese as the man-machine interaction language.

3.1. The architecture of the system

As the capability of storage and computation of a mobile device is limited, most of the data and knowledge should be stored by a server. The proposed system is realized by the Client/Server architecture. The Context-based Support Server Application (CSSA) realizes most of computations. The database at the server side stores the knowledge and the learning materials. The Context-based Support Mobile Application (CSMA) provides the man-machine interaction interfaces and displays the learning materials. The database at the client side only stores the user’s profile and the device’s profile, which may be used by the searching process for suitable language-learning materials. The CSMA is connected to the CSSA by the wireless network. The maintenance of the database and the
knowledge base will be carried out by language-teaching experts and knowledge engineers respectively at the server side, which isn’t the focus of the research.

3.2. The contexts related with mobile language-learning process

Dey defined context as follows. Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and application themselves. By improving the computer’s access to context, we can increase the richness of communication in human-computer interaction and make it possible to produce more useful computational services. As the aggregation of context is all about the whole situation relevant to an application and its set of users, we cannot enumerate which aspects of all situations are important, as this will change from situation to situation. For example, in some cases, the physical environment may be important, while in others it may be completely immaterial. Therefore, what kinds of context are important is decided by what the user wants to get from the application. In this research, the following aspects are important for realizing the functions of mobile language-learning support: the learner’s profile, the mobile device’s profile, the activity that the learner is doing or going to do, and the context related with the last language-learning process of the learner. Table 1 lists the specific context that each aspect should include.

Table 1. Specific Context.

<table>
<thead>
<tr>
<th>Context Aspect</th>
<th>Specific Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learner’s profile</td>
<td>LanguageLevel</td>
</tr>
<tr>
<td>Device’s profile</td>
<td>WirelessNetworkSpeed, MultimediaSupportKind</td>
</tr>
<tr>
<td>Activity</td>
<td>Query</td>
</tr>
<tr>
<td>Last language-learning process</td>
<td>Business, MultimediaSupportKind, LanguageLevel</td>
</tr>
</tbody>
</table>

Specifically, LanguageLevel stands for a user’s Chinese level. Only those learning materials that fit or near to the learner’s level should be supplied to him/her. According to the Chinese Proficiency Test (China’s Hanyu Shuiping Kaoshi, known as HSK), LanguageLevel may be Basic, Elementary-Intermediate, or Advanced. Device’s WirelessNetworkSpeed and MultimediaSupportKind are used to decide the learning materials with what kind of multimedia should be selected. WirelessNetworkSpeed will be categorized to four kinds according to the kinds of multimedia it suitable for transmitting. MultimediaSupportKind may be audio, video, image or text. Query is usually used to launch a language-learning process, it is related with what activity that the learner is doing or going to do. The activity or query should relate to a business, as learning materials are categorized by different businesses in different domains. Business is used to retrieve the materials related with it. Business, MultimediaSupportKind and LanguageLevel will be the ultimate context used to decide the most suitable language-learning materials, so the last language-learning context will only keep them for recovering the support process for the last learning.

Based on the ways of obtaining them, the above mentioned specific contexts can be categorized into three kinds, predefined context, historical context and dynamic context. Table 2 lists the category.

Table 2. Category of the Specific Context.

<table>
<thead>
<tr>
<th>Context Category</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predefined context</td>
<td>Learner’s profile(LanguageLevel), Device’s Profile(MultimediaSupportKind)</td>
</tr>
<tr>
<td>Historical context</td>
<td>Last language-learning process(Business, MultimediaSupportKind, LanguageLevel)</td>
</tr>
<tr>
<td>Dynamic context</td>
<td>Device’s Profile(WirelessNetworkSpeed), Activity(Query)</td>
</tr>
</tbody>
</table>

A learner can test his/her Chinese level by a self-help mock test at the website of http://www.hsk.org.cn/english/intro_sample.aspx, and input the level into the system as a predefined context. What kind of multimedia that a mobile device can support is decided by the device itself, which cannot be changed after the device was produced, so this context is a predefined one. Historical context reflected by the three aspects, business, multimedia support kind, and language level is used to recover the support for the last learning process. Predefined context and historical context are stored in the database of the CSMA, which can be obtained in real time.
by searching the database when they are needed. These two kinds of context are the static context. Dynamic context means that it will dynamically changed with the environment. Wireless network speed of a device is dynamically changed with the condition of the wireless network and the current environment under which the device is, which can be automatically sensed by the speed sensor of the mobile device, while the context of activity is dynamically changed with what the learner is doing or what he is going to do. However, the obtaining process of the context related to business and query is complicated, which needs the support of the representation for the domain knowledge. Additionally, how the context can be matched with suitable language-learning materials is determined by the knowledge representation method for context and learning materials. Hence, in next section, we’ll focus on the knowledge representation method.

3.3. Context-based ontology

Confirming what kind of business that the learner is doing or going to do is very important for selecting the learning materials related with the business. In order to realize this process, two kinds of ontology should be used, general ontology and domain ontology. Domain stands for the area that the business belongs to. For example, the business of “time deposit” belongs to the domain of Bank. The business of “order a dish” belongs to the domain of Restaurant. “Time deposit” or “order a dish” is a kind of business. As the domain ontology has limited concepts that may lead to no match between an inputted word and the domain ontology concept, the general ontology can complement it. During developing the system, we use WordNet ontology as the general ontology. WordNet® is a large lexical database of English, developed under the direction of George A. Miller (Emeritus). Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. The resulting network of meaningfully related words and concepts can be navigated with the browser. WordNet is also freely and publicly available for download. WordNet's structure makes it a useful tool for computational linguistics and natural language processing.

Besides the development of the domain ontology and the use of the WordNet ontology, the context should also be modelled based on ontology. The reasons for it could be found in reference. In order to support the matching process between context and language-learning materials, the ontology should include the necessary relations between them. By developing the domain ontology and the context model based on ontology, and combining the relations between context and language-learning materials, a context-based ontology is constructed, which is shown by Figure 1. This ontology consists of three sub-ontologies, Domain, Context and LanguageLearningMaterial. Domain defines the businesses that people may do in their daily life as well as the relationships among businesses. Context defines a learner’s (i.e. user’s) activity, profile, mobile device’s attributes and states, the last language learning process as well as their relationships. LanguageLearningMaterial defines the main properties of language learning materials. In order to provide a learner with language-learning materials related with what he is doing or what he is going to do, Business is set to be the core object that relates context and language learning materials in this ontology. The Context sub-ontology is related with Business by the object Query’s function “matchBusiness” and the object LastLanguageLearningProccesss’s function “containBusiness”. The LanguageLearningMaterial sub-ontology is related with Business by the object LanguageLearningMaterial’s function “containBusiness”. As this ontology could cover the objects, their properties and functions that would be used by inference rules in the matching process, for simplicity, it isn’t detailed enough to include data properties.

3.4. The matching process

The matching process between the context and language-learning materials will be realized by the support of the context-based ontology, WordNet ontology and the database used to store language-learning materials.

The general matching process can be illustrated by Figure 2. In this figure, the WordNet ontology will support the process of Step I, and the context-based ontology will support the process of step I, II, and III, while the database used to store language-learning materials will support all of the four steps. The output of step I is a set of learning materials, named B_LLM_SET, which is the input of step II, and the output of step II is a subset of B_LLM_SET, named B_U_LLM_SET, which is the input of step III, and the output of the step III is a subset of B_U_LLM_SET,
named B_U_D_LLM_SET, which is the input of step IV. In the following paragraphs, each step will be unfolded in detail.

Step I. Confirm the business and select language-learning materials by the business.

Figure 3 shows the process of confirming the business that the learner is doing or going to do. After the man-machine interaction process finishes, the business will be confirmed. Based on the confirmed business and the properties in context-based ontology, Rule 1 defined below matches the context of Activity with the suitable language-learning materials.

**Rule 1:**

\[
\text{suitableFor}(\text{LanguageLearningMaterial}, \text{Activity}) \quad \text{hasQuery}(\text{Activity}, \text{Query}) \quad \text{matchBusiness}(\text{Query}, \text{Business}) \quad \text{containBusiness}(\text{LanguageLearningMaterial}, \text{Business})
\]

The result of step I is a set of language-learning materials firstly selected by the process of confirming the business and Rule 1. We name it B_LLM_SET, which is the input of the next step.

Step II. Select language-learning materials from B_LLM_SET by the language level of the user.

As the user’s language level is a predefined context, when it is used in this step, it will be read directly from the database in CSMA. This step will select those learning materials with the same language level as the user’s language level from B_LLM_SET.

Step II is fulfilled by Rule 2 and Rule 3. Rule 2 defines what kinds of materials in set B_LLM_SET have the same language level with the user. The last item in this rule is a comparison built-in of SWRL, which will be satisfied if the first argument and the second argument are the same. Rule 3 matches the context of User with suitable language-learning materials in set B_LLM_SET. This rule will use the inference result of Rule 2.

Fig. 1. The Context-based Ontology.
Rule 2:
matchTo(?UserLanguageLevel, ?LanguageLearningMaterialLanguageLevel)
suitableFor(?LanguageLearningMaterial, ?Activity)
hasActivity(?User, ?Activity)  hasLanguageLevel(?User, ?UserLanguageLevel)
hasLanguageLevel(?LanguageLearningMaterial, ?LanguageLearningMaterialLanguageLevel)
swrlb:equal(?UserLanguageLevel, ?LanguageLearningMaterialLanguageLevel)

Fig. 2. The General Matching Process between the Context and Language-Learning Materials.

Fig. 3. The process of confirming the business that the learner is doing or going to do
Rule 3:
\[
\text{suitableFor}(\text{?LanguageLearningMaterial}, \text{?User}) \\
\text{hasLanguageLevel}(\text{?User}, \text{?UserLanguageLevel}) \\
\text{hasLanguageLevel}(\text{?LanguageLearningMaterial}, \text{?LanguageLearningMaterialLanguageLevel}) \\
\text{matchTo}(\text{?UserLanguageLevel}, \text{?LanguageLearningMaterialLanguageLevel})
\]

The result of step II is a subset of B\_LLM\_SET. We name it B\_U\_LLM\_SET. B\_U\_LLM\_SET is the input of the next step.

- **Step III. Select language-learning materials from B\_U\_LLM\_SET by the multimedia support kind suitable for current device and current wireless network speed.**

In this step, the first work should be determining what kind of multimedia support is suitable for the current wireless network and the device. Table 4 defines what kinds of multimedia are suitable to be provided to a device under different speeds of the wireless network.

<table>
<thead>
<tr>
<th>Wireless network speed</th>
<th>The suitable multimedia support kinds</th>
</tr>
</thead>
<tbody>
<tr>
<td>WirelessNetworkSpeed ( \leq ) 50Kbps</td>
<td>MultimediaSupportKind: {text}</td>
</tr>
<tr>
<td>50Kbps &lt; WirelessNetworkSpeed ( \leq ) 100Kbps</td>
<td>MultimediaSupportKind: {text, picture}</td>
</tr>
<tr>
<td>100Kbps &lt; WirelessNetworkSpeed ( \leq ) 150Kbps</td>
<td>MultimediaSupportKind: {text, picture, audio}</td>
</tr>
<tr>
<td>WirelessNetworkSpeed &gt; 150Kbps</td>
<td>MultimediaSupportKind: {text, picture, audio, video}</td>
</tr>
</tbody>
</table>

Rule 4 decides what kind of multimedia is suitable for the current device according to the current speed of wireless network and the multimedia kind supported by the device. The last item in this rule is a list built-in of SWRL, which will be satisfied if the first argument is a list containing elements found in both the list the second argument and the list the third argument. And based on the determined multimedia support kind, this rule finally matches the context of Device with the suitable language-learning materials in set B\_U\_LLM\_SET.

Rule 4:
\[
\text{suitableFor}(\text{?LanguageLearningMaterial}, \text{?Device}) \\
\text{suitableFor}(\text{?LanguageLearningMaterial}, \text{?Activity}) \\
\text{hasActivity}(\text{?User}, \text{?Activity}) \\
\text{suitableFor}(\text{?LanguageLearningMaterial}, \text{?User}) \\
\text{usedBy}(\text{?Device}, \text{?User}) \\
\text{hasWirelessNetworkSpeed}(\text{?Device}, \text{?WirelessNetworkSpeed}) \\
\text{suitableFor}(\text{?WirelessNetworkSpeed}, \text{?WMultimediaSupportKind}) \\
\text{hasMultimediaSupportKind}(\text{?Device}, \text{?DMultimediaSupportKind}) \\
\text{hasMultimediaSupportKind}(\text{?LanguageLearningMaterial}, \text{?LMultimediaSupportKind}) \\
\text{swrlb:listIntersection}(\text{?LMultimediaSupportKind}, \text{?WMultimediaSupportKind}, \text{?DMultimediaSupportKind})
\]

The result of step III is a subset of B\_U\_LLM\_SET. We name it B\_U\_D\_LLM\_SET. B\_U\_D\_LLM\_SET is the input of the next step.

- **Step IV. Rank the items in the set B\_U\_D\_LLM\_SET.**

As there may be several or more items in the set B\_U\_D\_LLM\_SET, which cannot be provided simultaneously to the learner. They should be provided one by one, or, the learner usually only needs one of them. Hence, ranking them in order to find the most suitable one is the work of this step.

The tf–idf weight (term frequency–inverse document frequency) is a weight often used in information retrieval and text mining. This weight is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. The importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the corpus. As variations of the tf–idf weighting scheme are often used by search engines as a central tool in scoring and ranking a document’s relevance given a user query, we adopt this technique to rank the language-learning materials in the set B\_U\_D\_LLM\_SET.
4. Implementation of the prototype system: an example of supporting mobile Chinese learning under a bank business

The system is implemented based on Client/Server structure taking advantage of the internet and .NET platform. With respect to GUI, C# programming language is used for the interface of man-machine interactions. The database in the server side is SQL server 2005 with its visualized tools. IIS is used for web services and ADO.NET for database access. Protégé 0.4 is used for building the ontologies. The parser of OWL ontologies is OWL API 2.2.0. A C# API and reference implementation are used for creating, manipulating and serializing OWL Ontologies. Ontology reasoning has been done in FaCT++, where the FaCT++ is one of reasoning tools with complete OWL-DL consistency checker and supports standard set of Description Logic inference services. The whole prototype system is implemented in the platform of Visual studio.net 2008 for development and test.

The scenario is that Mr. Lee from America is now in a Chinese bank for depositing some money. He wants to know the dialogue about this kind of business between a customer and a cashier. Assume that Mr. Lee input the query by the word “deposit”. Then, this word will be sent to the Context-based Support Server Application (CSSA). According to the process shown by Figure 3, the domain concepts in context-based ontology stored in the knowledge base of CSSA will be firstly tried. As “deposit” cannot be matched with a domain concept, the WordNet ontology will return the synonyms of “deposit” for Lee to choose the one with the nearest meaning, see Figure 4(a). After Lee chooses “bank deposit”, the domain of “bank” will be matched successfully according to the branch of domain in the context-based ontology. Then, the businesses related with the domain of “bank” will be shown to him, see Figure 4(b). After the business of “deposit in personal business” is selected, the sub-businesses of it will be listed, see Figure 4(c).

5. Conclusions and future work

This paper presents a prototype context-based support system of mobile Chinese learning for foreigners in China. By the context-based ontology for representing the knowledge, the system can match the context of a learner with
suitable Chinese-learning materials. It cannot only support the Chinese-learning process at any time and at any place by providing a learner with learning materials, but also can provide him/her with suitable learning materials related with the contexts of what he/she is doing or going to do, his/her profile and mobile device. Therefore, it greatly facilitates the Chinese-learning process and improves the effect and efficiency of the learning.

The proposed system has a great potential market, as China is becoming one of the largest mobile business markets for its great amount of users of mobile devices, and more and more foreigners are coming to China who may need this kind of system. Moreover, the system will also facilitate the diffusion of the Chinese culture, as it has an additional benefit that a foreigner can experience or learn more about Chinese culture by learning Chinese under real-world surroundings. As we know, language is a media and also a carrier of the culture. As the system can support Chinese learning under the real-world communication environment, the deeper cultural implications in Chinese language can be experienced in daily life by foreign learners.

Acknowledgements

This work has been partially supported by the grants from Natural Science Foundation of China (Nos. 71201014, 71272093, 71271037, 71171029); The Fundamental Research Funds for the Central Universities (Nos. DUT14QY28, DUT14RC(4)04)

References