This article presents a bilingual ontology-based dialog system with multiple services. An ontology-alignment algorithm is proposed to integrate ontologies of different languages for cross-language applications. A domain-specific ontology is further extracted from the bilingual ontology using an island-driven algorithm and a domain corpus. This study extracts the semantic words/concepts using latent semantic analysis (LSA). Based on the extracted semantic words and the domain ontology, a partial pattern tree is constructed to model the speech act of a spoken utterance. The partial pattern tree is used to deal with the ill-formed sentence problem in a spoken-dialog system. Concept expansion based on domain ontology is also adopted to improve system performance. For performance evaluation, a medical dialog system with multiple services, including registration information, clinic information, and FAQ information, is implemented. Four performance measures were used separately for evaluation. The speech act identification rate was 86.2%. A task success rate of 77% was obtained. The contextual appropriateness of the system response was 78.5%. Finally, the rate for correct FAQ retrieval was 82%, an improvement of 15% over the keyword-based vector-space model. The results show the proposed ontology-based speech-act identification is effective for dialog management.

Introduction

Conversation with the machine using spoken language is a tremendous vision of human machine interface (Lee, Jian, & Huang, 2005; Yuan & Sun, 2005). Dialog management is arguably the core functionality of spoken-dialog systems, because the dialog manager tends to perform the controlling function of the system as a whole, and maintains a model of the evolving dialog context in a ubiquitous computing environment (Huang, Acero, & Hon, 2001; Tombros, & Crestani, 2000). However, useful knowledge representation is vital for human-machine interfaces. As global communication and multietnic societies grow, the demand for multilingual capability will increase. In a multilingual spoken-dialogue application, input speech is sometimes spoken in two languages, especially in certain domains. In recent decades, several practicable dialog systems, such as air-travel information-service systems, weather forecast systems, automatic banking systems, automatic train timetable information systems, and the Circuit-Fix-it shop system, have been developed. The dialog control in these systems can be categorized into three different approaches: finite state-based, frame-based, and agent-based (McTear, 2002). With these approaches, there are two major issues in dialog-management research; the first is the robustness for speech recognition and the other is speech-act identification. Since speech act plays an important role in the development of dialog management for dealing with complex applications, speech-act identification (Wu, Yan, & Lin, 2002) becomes the most important topic with respect to the methods used to control the dialog with the users.

Although spontaneous spoken-language understanding is an open issue in natural-language processing and artificial intelligence (Smith, 1998; Ward & Tsukahara, 2003; Potamianos, Narayanan, & Riccardi, 2005), in practice, the concept of understanding is situation dependent and domain specific. The ontology, a specification of a conceptualization in a specific domain, is used as the knowledge representation to provide a quantitative approach for linguistic application, especially in understanding (Sarikaya, Gao, Picheny, & Erdogan, 2005). For example, Lee et al. (2005) proposed a fuzzy ontology and applied it to news summarization (Roy, Hsiao, & MAVridis, 2004). Yuan and Sun developed a mobile query system using latent semantic analysis via
Dialog systems usually collect information from users by filling semantic frames or slots. There are many strategies to obtain the content desired in the semantic frames/slots. The plan-based approaches adapt to the user’s interest to gather the information items. The object-oriented approaches inherit the attribute from the ancestor classes. However, neither the plan-based nor the object-oriented approaches work effectively. This is because these approaches cannot identify the precise speech act from the user’s utterances using only semantic frames or slots. In addition, more accurate identification is needed for applications offering multiple services in order to distinguish different speech acts of utterances with different word orders. This study proposes an approach to identify the speech act in a complex dialog situation such as multiple conversational interfaces (Yuan & Sun, 2005). Recently, significant effort has focused on constructing ontologies manually according to domain experts’ knowledge. However, manual ontology merging using conventional editing tools without intelligent support is difficult, labor intensive, and error-prone. This study employs an ontology-alignment algorithm (Yeh, Wu, Chen, & Yu, 2004) to generate a new ontology by integrating the knowledge from aligned authoritative ontologies. Based on the integrated ontology, we can interpret the user’s utterances more precisely. Several systems and frameworks to help knowledge engineers perform ontology merging have recently been proposed (Noy & Musen, 2003; Gennari et al., 2003). To avoid reiteration in ontology construction, algorithms for ontology merging (UMLS http://umlsks.nlm.nih.gov; Langkilde & Knight, 1998) and ontology alignment (Vossen, Diez-Orzas, & Peters, 1997; Weigard & Hoppenbrouwers, 1998; Asanoma, 2001) have been investigated. In these approaches, the final ontology is a merged version of the original ontologies with aligned links between them (Daudé, Padró, & Rigau, 2003). Automatic and semiautomatic methods have been developed. OntoExtract (Fensel et al., 2002; Missikoff, Navigli, & Velardi, 2002) provides an ontology-engineering chain for constructing domain ontology from WordNet and SemCor. Some recent approaches have been discussed in Ezuenat et al. (2003). The alignment approaches were classified as local or global methods. Four main local methods, the terminological, extensional, semantic, and structural methods, were introduced to measure the correspondence between two ontologies at the local level. Yang and Li proposed an approach to align the Chinese and English corpus automatically using lexical and grammatical features (Yang & Li 2003). Nowadays, much work is being invested in ontology construction for domain applications. Performing authoritative evaluation of ontologies is becoming a critical issue. Some evaluation methods were integrated into ontology tools to detect and prevent mistakes that might be made in the course of developing taxonomies with frames, as described in Gómez-Pérez (2001). These approaches defined three main types of mistakes: inconsistency, incompleteness, and redundancy mistakes. In this approach, the system starts from a small ontology kernel and constructs the ontology through text-understanding automatically.

Discourse information is generally stored in the semantic frame or slot for speech-act identification. However, dialog systems cannot obtain the exact semantic meaning due to the lack of knowledge or conceptual representation in the input utterances. It is manifest that words are the semantic building blocks in linguistic research, so the conceptual representation of a word can provide auxiliary aid for the comprehension of input utterances. This study introduces an ontology that provides the conceptual space for domain application. Therefore, key-concept expansion-covering synonyms and relations defined in the ontology are used for semantic inference (Staab, Studer, Schnurr, & Sure, 2001).

The rest of this article is organized as follows. We present the framework for ontology-based dialog management for a conversational human-machine system. Then, ontology alignment and domain ontology extraction are introduced in detail. The semantic word collection and partial pattern tree bank are described. Next, we define the ontology-based dialog system with speech-act identification using the partial pattern tree. Experimental results and discussions are presented. Finally, we summarize the findings and describe future work.

**Framework of the Ontology-Based Dialog System**

As shown in Figure 1, the conversational-dialog system for multiple services in the medical domain is composed of three main components: semantic words extraction using latent semantic analysis (LSA), speech-act identification based on ontology and partial pattern trees (PPTs), and template-based response generation for speech synthesis.

In this system, the user’s speech input is fed into the speech recognizer to obtain the word lattice. According to the pretrained partial pattern trees and the integrated ontology, a speech act is identified from the concept and the structural information of the recognized words. The semantic filler extracts the concepts to fill in the semantic slots, and the dialog manager generates the response utterance according to the status of the filled semantic slots. Finally, a template-based response-generation mechanism is used to provide an appropriate response to the user.

There are three medical services contained in this study: a medical Frequently Asked Questions (FAQ) information
service, a clinic-information service, and a registration-information service. The medical Frequently Asked Questions (FAQ) information service provides the function of mining the health-education document in the FAQ form. The clinic-information service provides the function of determining the clinic department in which the symptoms provided by users can be treated. The registration-information service provides online registration for making an appointment with the doctor using the conversational interface.

Ontology Alignment and Extraction for Domain Application

In this study, a bilingual ontology is constructed by aligning Chinese words in HowNet (http://www.keenage.com/) with their corresponding synsets defined in WordNet (http://www.cogsci.princeton.edu/~wn/) according to the co-occurrence of the words in a bilingual parallel corpus as shown in Figure 2. In HowNet, each lexical entry is defined as a combination of one or more primary features and a sequence of secondary features. The primary features indicate the entry’s category, for example, the relation “is-a” in a hierarchical structure. Overall, 1,521 primary features are divided into six upper categories: event, entity, attribute value, quantity, and quantity value. These primary features are organized as a hierarchical structure. In WordNet, a word may be associated with many synsets, each corresponding to a different sense of the word. To find a relation between two different words, all the synsets associated with each word are considered.

For alignment, the parallel corpus Sinorama (Sinorama Magazine and Wordpedia Co., 2001) database, containing over 6,500 documents with 48,000,000 words collected from 1976 to 2000 in Chinese and English, is adopted as the training corpus to compute the conditional probability of the words in WordNet, given the words in HowNet. The goal of this approach is to increase the amount of relation and structural information coverage by aligning semantic concepts in WordNet and HowNet.

Equation 1 shows the alignment between the words in HowNet and the synsets in WordNet. Given a Chinese word, \( CW_i \), the probability of the word being related to synset \( \text{synset}^k \), can be obtained via its corresponding English synonyms, \( EW_j^k, j = 1, \ldots, m \), which are the elements in \( \text{synset}^k \). The probability is estimated as follows:

\[
\Pr(\text{synset}^k | CW_i) = \frac{1}{m} \sum_{j=1}^{m} \Pr(\text{synset}^k, EW_j^k | CW_i) \tag{1}
\]

where

\[
\Pr(\text{synset}^k | EW_j^k, CW_i) = \frac{N(\text{synset}^k, EW_j^k, CW_i)}{\sum_i N(\text{synset}^k, EW_j^k, CW_i)} \tag{2}
\]
In the above equation, \( N(\text{synset}^j, \text{EW}^j, \text{CW}) \) represents the number of co-occurrences of \( \text{CW} \), \( \text{EW}^j \), and \( \text{synset}^j \). The probability \( \Pr(\text{EW}^j \mid \text{CW}) \) is set to one when at least one of the primary features, \( PF_i^j(\text{CW}) \), of the Chinese word \( \text{CW} \) defined in HowNet matches one of the ancestor nodes of \( \text{synset}^j \), excluding the root nodes in the hierarchical structures of the noun and verb. Otherwise, the probability \( \Pr(\text{EW}^j \mid \text{CW}) \) is set to zero:

\[
\Pr(\text{EW}^j \mid \text{CW}) = \begin{cases} 
1, & \text{if } (\bigcup PF_i^j(\text{CW}) - \{\text{entity, event, act, play}\}) \cap (\bigcup \text{ancestor}(\bigcup \text{synset}^j(\text{EW})) - \{\text{entity, event, act, play}\}) \neq \emptyset, \\
0, & \text{Otherwise} 
\end{cases} 
\] (3)

For extending the ontology to domain applications, more detailed definitions and terminology are required. This article proposes a two-stage domain ontology-extraction method. This approach extracts the domain ontology from the cross-language ontology by using the island-driven algorithm in the first stage. The terminology and axioms defined in a medical encyclopedia are integrated into the domain ontology in the second stage.

Generally, an ontology provides consistent concepts and word representations necessary for clear communication within the knowledge domain. Even in domain-specific applications, the number of words can be expected to be huge. Synonym pruning is an effective way to perform word-sense disambiguation. This article proposes a corpus-based statistical approach to extracting a domain ontology. The steps are as follows:

**Step 1. Linearization.** In this step, the tree structure in the general-purpose ontology is decomposed into a vertex list that is an ordered node sequence starting at the root node and ending at the leaf nodes.

**Step 2. Concept extraction from the corpus.** The node is defined as an operative node when the tf-idf value of word \( W_i \) in the domain corpus is higher than that in its corresponding contrastive (out-of-domain) corpus. That is,

\[
\text{operative_node}(W_i) = \begin{cases} 
1, & \text{if } tf - \text{idf}_{\text{domain}}(W_i) > tf - \text{idf}_{\text{contrastive}}(W_i) \\
0, & \text{Otherwise} 
\end{cases} 
\] (4)

where

\[
tf - \text{idf}_{\text{domain}}(W_i) = \text{freq}_{W_i, \text{Domain}} \times \log \frac{n_i, \text{Domain} + n_i, \text{Contrastive}}{n_i, \text{Domain}},
\]

\[
tf - \text{idf}_{\text{contrastive}}(W_i) = \text{freq}_{W_i, \text{Contrastive}} \times \log \frac{n_i, \text{Domain} + n_i, \text{Contrastive}}{n_i, \text{Contrastive}}.
\]

In the above equations, \( \text{freq}_{W_i, \text{Domain}} \) and \( \text{freq}_{W_i, \text{Contrastive}} \) are the frequencies of word \( W_i \) in the domain documents and its contrastive (out-of-domain) documents, respectively; \( n_i, \text{Domain} \) and \( n_i, \text{Contrastive} \) are the numbers of documents containing word \( W_i \) in the domain documents and its contrastive documents, respectively.

**Step 3. Relation expansion using the island-driven algorithm.** Some domain concepts are no longer operative after the previous steps are performed, due to the problem of data sparseness. According to the analysis performed during ontology construction, most of the inoperative concept nodes have operative hypernym nodes and hyponym nodes. Therefore, the island-driven algorithm is adopted to activate these inoperative concept nodes if their ancestors and descendants are all operative.

**Step 4. Domain ontology extraction.** In the final step, the linear vertex list sequence is merged into a hierarchical tree. However, some noisy concepts defined as nodes not belonging to this domain are operative according to Equation 4.
Accordingly, the second goal is to filter out the nodes with operative noisy concepts. In this step, noisy concepts with no ancestors or descendants belonging to the domain are removed. Finally, the domain ontology is extracted.

In practice, specific domain terminology and axioms should be derived and introduced into an ontology for domain-specific applications (Abu-Hanna, Corneta, de Keizer, Crubézy, & Tu, 2005). There are two approaches to integrating terminology and axioms into an ontology: the first one is manual editing performed by ontology engineers, and the second is automatic integration from a domain encyclopedia. We derived 1,213 axioms from a medical encyclopedia with terminology related to diseases, syndromes, and clinic information (Wu, Yeh, & Chen, 2005).

Semantic Word Collection and Partial Pattern Tree Construction

A complicated issue in our study is the cross-language usage of utterances as described in (Wu, Chiu, Shia & Lin 2006). This situation arises very commonly in Asia. In Taiwan, three languages, Chinese, Taiwanese, and English, are frequently used and spoken in medical conversations; in particular, English names entities and English terminologies are used even when English is not the primary language of the conversation. Since the sentence patterns and terminologies are very similar between Chinese and Taiwanese, we can regard the sentence pattern and terminology of Taiwanese as those of Chinese. For cross-language application in dialog systems, there are four pattern and terminology of Chinese/Taiwanese native speakers, for example: {I have got a headache, a stomachache, and a fever.}

1. **Chinese sentence pattern with Chinese terminologies.** This category, which forms the overwhelming majority in our collection, is used by Chinese/Taiwanese native speakers in Taiwan, for example: {我感到頭痛，胃痛，發燒等症狀} (I have got a headache, a stomachache, and a fever.)

2. **Chinese sentence pattern with English terminologies.** The utterances belonging to this category have progressively increased recently, especially in the language usage of medical personnel, for example: {我有 headache, stomachache, fever 等症狀} (I have got a headache, a stomachache, and a fever.)

3. **English sentence pattern with English terminologies.** The utterances in this category are usually used by non-Chinese/Taiwanese native speakers, for example: {I have got a headache, a stomachache, and a fever.}

4. **English sentence pattern with Chinese terminologies.** The utterances in this category are also used by non-Chinese/Taiwanese native speakers. There are two major reasons to form this utterance: the first one is the speaker’s desire to prevent the non-English native listener from misunderstanding, for example: {I have got a 頭痛，胃痛，發燒}. (I have got a headache, a stomachache, and a fever.) The second reason for these utterances is to discuss Chinese-named entities or the terminologies defined in traditional Chinese medicine, such as {氣虛} (deficiency of vital energy). However, only a few utterances belong to this category in our collection.

Since the partial pattern trees are constructed by sentence patterns, the aligned bilingual ontology is employed to address the problem of sentence patterns and terminologies using different languages within the same utterance. In fact, we translate the terminologies from the language usage of terminology to that of sentence patterns based on the bilingual ontology constructed as discussed in the previous section.

In speech-act identification, word matching is not suitable for spontaneous speech due to the difference between the word orders in two spoken sentences. Consequently, in this study, a partial pattern tree (PPT) is used to partially match the words between the ill-formed input utterance and the speech-act patterns in the PPT for robust speech-act identification. Two preprocesses, corpus collection and partial pattern tree bank construction, are essential for ontology-based speech-act identification.

**Semantic Words Extraction Based on Latent Semantic Analysis**

In statistical approaches, the corpus generally plays an important role in parameter estimation and model construction. There are three methods for corpus collection: telephone recording and transcription, wizard of Oz (WOZ), and online collection, shown in Figure 3. Telephone recording and transcription means collecting the data by recording the service phone call between humans for transcription. In the wizard of Oz (WOZ) method, the dialogs between the subject and the system, monitored by a human via internet, are recorded. In this method, the users think they are talking with an automatic dialog system. After the prototype of the system has been established, the online corpus can be collected via the real online dialog between users and the automatic dialog system.

Latent semantic analysis is a novel approach for automated document indexing that is based on a latent class model for factor analysis of count data. There are two reasons our approach uses latent semantic analysis. First, semantic factors extracted from latent semantic analysis are more suitable for speech-act representation than keywords selected by human labor. Second, latent semantic analysis, which projects higher dimensions into the latent semantic space from the original space, provides a method for dimensionality reduction. This study constructs a SpeechAct-by-Word matrix first. The mapping is performed by decomposing the SpeechAct-by-Word matrix $A$ into the product of three matrices, $W$, $S$, and $SA$ using singular value decomposition:

$$A_{t \times d} = W_{t \times n} S_{n \times n} (SA_{d \times n})^T \approx W_{t \times r} S_{r \times r} (SA_{d \times r})^T = \tilde{A}_{t \times d}$$ (5)

where $n = \min(t, d)$. The matrices $W$ and SA have orthonormal columns. This means that column vectors have unit length and all vectors are orthogonal to each other. In fact, these vectors are the eigenvectors. The diagonal matrix $S$ contains the singular values of $A$ in descending order. The $i$-th singular value indicates the amount of variations along
the $i$-th axis. The LSA approximation of $A$ is computed by thresholding all but the largest $r$ singular value in $S$. Besides dimensionality reduction, the LSA keeps the semantic words while filtering out the irrelevant words.

Construction of Partial Pattern Tree

In general, each utterance can be represented as a sequence of function phrases and a semantic word from the keyword set as follows:

$$S_i = \{FP_{i1}, FP_{i2}, \ldots, FP_{iN_i}, SW_i, FP_{iN_i+1}, \ldots, FP_{iN_{i+1}}\}$$

where $SW_i$ denotes the semantic word and $FP_j$ denotes the $j$-th function word in the utterance $S_i$. $N_{Bi}$ and $N_{Ai}$ represent the number of function words before and after the semantic word, respectively. Therefore, $2^{N_{Bi}+N_{Ai}}$ partial pattern sequences containing the semantic word, $SW_i$, will be generated according to the definition of PPT. As illustrated in Figure 4, the training sentence “I need a diagnosis” with a semantic word “diagnosis” will generate eight partial pattern sequences, such as “I need a diagnosis,” “I need … diagnosis,” “I … a diagnosis,” “I … diagnosis,” “… need a diagnosis,” “… need … diagnosis,” “… a diagnosis” and “… diagnosis.” Each partial pattern is presented as a path in a partial pattern tree from the root node to the leaf node, denoting the corresponding pattern, as shown in Table 1. For practical implementation, the words in the sequence without the semantic word are defined as the “don’t care function words” according to the position. Patterns are formed by neglecting the function words with “don’t care function words” bits labeled 1. However, the robustness of variance in these positions is kept for similar utterances with the same partial pattern.

The PPT is basically an integrated tree structure of the partial pattern sequences generated from the training sentences. Each partial pattern sequence is tagged with one speech act. Each internal node representing a word in the partial pattern tree is denoted as $IN_i = \{PH_i, FR_i, N_{Si}, Son_i\}$, where $PH_i$ is the word in this internal node. $FR_i$ is the frequency of the internal node. $N_{Si}$ is the number of descending internal nodes. $Son_i$ is the pointer linked to its son. Each external node represents one partial pattern sequence, which corresponds to one speech act in the PPT. The data structure of the external node is defined as $EN_i = \{PP_i, Ptr_i, P(SA_{ik})\}$, where $PP_i$ is the reference partial pattern sequence. $Ptr_i$ is the pattern pointer set. $P(SA_{ik})$ is the probability of the $k$-th speech act with respect to the $i$-th partial pattern sequence. The algorithm for constructing the partial pattern tree is described as follows:

PPT Construction Algorithm

1. Initialization: Create a root node, $R$.
2. Recursion: For each training partial pattern sequence $PP_i = \{PH_{i1}, PH_{i2}, \ldots, PH_{iN_{i}}\}$ where $N_{i}$ is the number of words in $PP_i$, compute Steps 2.1 to Step 2.5.
3.1. According to the sequence of words in the partial pattern sequence, traverse the PPT and stop at $IN_N$, such that the path from the root node to node $IN_N$ matches the prefix of $PP_i$.
3.2: If all the words in $PP_i$ have been used to traverse the PPT, then go to Step 2.5.
3.3: According to the suffix phrase $PH_{i}$ of $PP_i$, a subtree, which links $PH_{i}$, is created and the pointer of this subtree is added to the internal node $IN_N$.
   3.3.1: Create a new internal node, $IN_N$.
   3.3.2: $PH_i$ is the word sequence of $IN_N$.
   3.3.3: Add $IN_N$ into the son’s linked list of $IN_i$.
   3.3.4: Set $IN_N$ to $IN_i$.
3.4: If the external node does not exist in the partial pattern tree, then create a new external node $NE_N$ and set the word sequence $NE_N$ to the partial pattern $PP_i$.
3.5: Increase the frequency $FR$ of the internal node.
Speech Act Identification Using Ontology-based PPT

Using the constructed PPT, speech-act identification can be performed by matching the input utterance and the partial pattern sequence in the PPT. A dynamic programming algorithm is applied to deal with the problem of word order, which is important to the user’s intention. If the input utterance is $P = \{a_1, a_2, \ldots, a_I\}$ and $PP = \{b_1, b_2, \ldots, b_J\}$ is the external node representing the $j$-th partial pattern sequence, the similarity of these two sequences can be obtained using the Needleman-Wunsch algorithm (Needleman & Wunsch, 1970) shown in the following steps:

**Initialization.** Create a matrix with $I + 1$ columns and $J + 1$ rows. The first row and first column of the matrix can initially be filled with 0. That is,

$$\text{Sim}(i, j) = 0, \quad \text{if } i = 0 \text{ or } j = 0$$

**TABLE 1.** The illustration of partial patterns derived from the utterance “I need a diagnosis” with the semantic word “diagnosis.”

<table>
<thead>
<tr>
<th>Don’t care function words</th>
<th>b</th>
<th>b</th>
<th>b</th>
<th>Partial pattern</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>I need a</td>
<td>Original sentence.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>diagnosis.</td>
<td></td>
</tr>
<tr>
<td>Pattern-0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pattern-1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>I need</td>
<td>Neglecting the function word between “need” and “diagnosis.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>. . . diagnosis.</td>
<td></td>
</tr>
<tr>
<td>Pattern-2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>I . . .</td>
<td>Neglecting the function word between “I” and “a.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>diagnosis.</td>
<td></td>
</tr>
<tr>
<td>Pattern-3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>I . . . .</td>
<td>Neglecting the function words between “I” and “diagnosis.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>diagnosis.</td>
<td></td>
</tr>
<tr>
<td>Pattern-4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>. . . need</td>
<td>Neglecting the first function word.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>a diagnosis.</td>
<td></td>
</tr>
<tr>
<td>Pattern-5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>. . . . .</td>
<td>Neglecting the first words and the function words between “need” and “diagnosis.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>diagnosis.</td>
<td></td>
</tr>
<tr>
<td>Pattern-6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>. . . . .</td>
<td>Neglecting the two function words.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>diagnosis.</td>
<td></td>
</tr>
<tr>
<td>Pattern-7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>. . . . . .</td>
<td>Neglecting all function words.</td>
</tr>
</tbody>
</table>
Score pathways through array. Assign the values to the remnant elements in the matrix as the following:

\[
sim(i, j) = \max \left\{ \begin{array}{ll}
\text{Sim}(i - 1, j - 1) + \text{Sim}_{\text{onto}}(a_{i-1}, b_{j-1}), \\
\text{Sim}(i - 1, j) + \text{Sim}_{\text{onto}}(a_{i-1}, b_j), \\
\text{Sim}(i, j - 1) + \text{Sim}_{\text{onto}}(a_i, b_{j-1})
\end{array} \right. \]

Construct alignment. Determine the actual alignment with the maximum score \(\text{Sim}(P_i, PP_j)\).

It is difficult to obtain exact matching between the input utterance and the partial pattern sequence due to the versatility of spoken language, especial in word sense. To solve this problem, an ontology is employed to determine a word similarity measure. Two basic relations, hypernym and synonym, are introduced. The similarity is defined as follows:

\[
\text{Sim}_{\text{onto}}(a_i, b_j) = \begin{cases} 
1 & \text{if } a_i = b_j \\
\left(\frac{1}{2}\right)^l & \text{if } a_i \text{ and } b_j \text{ are hypernyms} \\
1 - \left(\frac{1}{2}\right)^n & \text{if } a_i \text{ and } b_j \text{ are synonyms} \\
0 & \text{otherwise}
\end{cases}
\]

where \(l\) is the number of levels between \(a_i\) and \(b_j\). The variable \(n\) is the number of their common synonyms in the synonym set. Finally, the speech act for the input utterance \(P_i\) is determined by the following equation.

\[
\text{SA}^*(P_i) = \arg \max_{k,j} \left\{ P(\text{SA}_k^j) \times \text{Sim}(P_i, PP_j) \right\} \quad (9)
\]

where \(P(\text{SA}_k^j)\) is the probability of the \(k\)-th speech act with respect to the \(j\)-th partial pattern sequence estimated in the construction of partial pattern trees.

Experiments

To evaluate the proposed method for ontology-based speech-act identification, a dialog system for medical application was developed as shown in Figure 5. There are three services provided in the medical domain with speech and text interfaces. The registration-information service, established based on the online registration system in National Cheng Kung University Hospital (http://140.116.253.119:2783/tandem/index.htm), helps patients register and consult the doctor’s contact information. The clinic-information service is a decision-support-oriented application system, which helps determine what clinic department should treat a patient with given symptoms by the inference axioms defined in the ontology. The FAQ information service is a health-care education system. It provides appropriate medical documents according
questions asked by the users (Simpson & Fraser, 2001). The speech acts in these three services are very diverse and manually classified into 12 speech acts as shown in Figure 6.

**Development of Spontaneous Corpus**

There were two corpora collected for our experiments: a telephone corpus and a wizard of Oz (WOZ) corpus. In total, 4,098 turns in 364 dialogs were recorded via the telephone line from National Cheng Kung University Hospital, Taiwan. There are 11.2 turns in each dialog, on average. The number of turns of most of dialogs range from 6 to 14. However, the long dialogs have a significant influence on the statistics. In this collection, the longest one contains 47 turns. We recorded and transcribed these dialogs as the training corpus.

We collected 836 turns in 90 dialogs via WOZ as the test corpus. WOZ 1 and WOZ 2 denote the corpus collected in the first and second sessions. We note that the number of turns using WOZ is smaller than that using phone-call recording. This is because the response of a computer is not natural enough to encourage users to talk with the machine. The distribution of the number of turns in a dialog for each corpus is shown in Figure 7. We find that there are two peaks in the distribution.

**Dialog Control Evaluation**

To evaluate the dialog-flow control using the proposed method, fifty individuals who did not participate in this research project were asked to test the system. There were four measures used for evaluation (Gibbon, Mertins, & Moore, 2000): task success rate, average number of turns, contextual appropriateness, and speech-act detection rate.

An approach using semantic frames was implemented for comparison; the results are listed in Table 2. According to these results, we can see the variability in the average number of turns. The number of turns in the FAQ information service is shorter than in the multiple-information services. The FAQ service achieves the best score in speech-act identification rate due to the clarity of its behavior compared with other services. However, we find that there is some confusion between the registration-information service and the clinic-information service in speech-act identification. The task success rate of the FAQ service is higher than that

**FIG. 7.** Distributions of the number of turns in a dialogue for three corpora.

**TABLE 2.** Evaluation measures for the conventional dialog system with semantic frame.

<table>
<thead>
<tr>
<th>Evaluation measures</th>
<th>Speech act identification rate (%)</th>
<th>Task success rate (%)</th>
<th>Average number of turns (%)</th>
<th>Contextual appropriateness (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration module</td>
<td>87.3</td>
<td>92</td>
<td>8.40</td>
<td>82.0</td>
</tr>
<tr>
<td>Clinic-query module</td>
<td>84.6</td>
<td>80</td>
<td>9.27</td>
<td>75.1</td>
</tr>
<tr>
<td>FAQ module</td>
<td>92.4</td>
<td>88</td>
<td>4.80</td>
<td>85.2</td>
</tr>
<tr>
<td>Integrated system</td>
<td>80.6</td>
<td>68</td>
<td>9.80</td>
<td>74.3</td>
</tr>
</tbody>
</table>
of other services. However, a dramatic decrease is seen in the integrated system due to the poor performance in speech-act identification. Speech-act identification is strongly related to a contextually appropriate response. In addition, Table 3 shows that the ontology-based partial pattern tree can effectively improve the speech-act identification rate and task success rate, especially with multiple services in a dialog system.

**Concept Extraction Evaluation**

One of the essential issues in achieving a friendly human-machine interface is semantic understanding, especially for a machine to capture the semantic concepts that humans desire to express. The concept-extraction mechanisms convert the human’s spoken language into a computer-readable semantic representation that is easier for programs to manipulate. The semantic representation of the proposed dialog system is the semantic frame (Minsky, 1975). The scenario metrics used in MUC-7 (Message Understanding Conferences; MUC-7, 2006) were adopted for assessing concept extraction, which is defined as follows:

\[
Recall = \frac{N_{COR}}{N_{COR} + N_{MIS}}, \quad (10)
\]

\[
Precision = \frac{N_{COR}}{N_{COR} + N_{SPU}}, \quad (11)
\]

\[
Undergeneration = \frac{N_{MIS}}{N_{COR} + N_{MIS}}, \quad (12)
\]

\[
Overgeneration = \frac{N_{SPU}}{N_{COR} + N_{SPU}} \quad (13)
\]

where \(N_{COR}\) denotes the number of concepts declared relevant when relevant is correct, and \(N_{SPU}\) is the number of concepts declared relevant when irrelevant is correct. \(N_{MIS}\) means the number of concepts declared irrelevant when relevant is correct. Table 4 shows the results of concept extraction. The point to observe is that the speech-act identification rate is essential for the concept-extraction task. Additionally, the improvement of concept extraction can achieve more friendly human-machine interface, especially in the average number of turns and contextual appropriateness.

**Conclusions and Future Work**

This article presents an approach for speech-act identification using an ontology-based partial pattern tree. In this study, an ontology alignment and domain ontology extraction algorithms using an island-driven algorithm are proposed. These approaches were applied to a complex dialog application: a system with multiple services in the medical domain. According to the experiments, the speech-act identification rate was 86.2%. A task success rate of 77% was obtained. The contextual appropriateness of the system response was 78.5%. Finally, the correct rate for FAQ retrieval was 82%, an improvement of 15% over the keyword-based vector-space model. Ontology can effectively capture meaning quantitatively in speech-act identification. The proposed method, considering the structural information represented by partial pattern trees, outperforms the traditional approach using the keywords. The results show the ontology-based partial pattern tree is superior to traditional dialog management, especially in precision and robustness. Future directions include constructing a semantic discourse model and implementing automatic online determination of the partial patterns to update the partial pattern tree bank. For these goals, novel algorithms to prevent errors from speech recognition should be developed.

**References**


---

**TABLE 3.** Evaluation measures for the dialog system with ontology-based partial pattern tree.

<table>
<thead>
<tr>
<th>Evaluation measures</th>
<th>Speech act identification rate (%)</th>
<th>Task success rate (%)</th>
<th>Average number of turns</th>
<th>Contextual appropriateness (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration module</td>
<td>94.2</td>
<td>95</td>
<td>7.30</td>
<td>85.0</td>
</tr>
<tr>
<td>Clinic-query module</td>
<td>90.4</td>
<td>82</td>
<td>8.20</td>
<td>79.4</td>
</tr>
<tr>
<td>FAQ module</td>
<td>94.1</td>
<td>92</td>
<td>4.70</td>
<td>88.8</td>
</tr>
<tr>
<td>Integrated system</td>
<td>86.2</td>
<td>77</td>
<td>9.20</td>
<td>78.5</td>
</tr>
</tbody>
</table>

**TABLE 4.** Results of concept-extraction evaluation.

<table>
<thead>
<tr>
<th></th>
<th>Recall</th>
<th>Precision</th>
<th>Undergeneration</th>
<th>Overgeneration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional approach</td>
<td>0.81</td>
<td>0.91</td>
<td>0.19</td>
<td>0.09</td>
</tr>
<tr>
<td>Ontology-based PPT</td>
<td>0.88</td>
<td>0.91</td>
<td>0.12</td>
<td>0.10</td>
</tr>
</tbody>
</table>


