An Intelligent Semantic Agent for Supervising Chat Rooms in e-Learning System

Ying-Hong Wang Wen-Nan Wang Chu-Chi Huang
Department of Computer Science and Information Engineering
Tamkang University, Tamsui, Taipei 251, Taiwan
inhon@mail.tku.edu.tw 892190033@s92.tku.edu.tw 793190090@s93.tku.edu.tw

Abstract

This paper presents an English chat room system in which students can discuss course by interacting with teachers and students. First, the novel mechanism provides Learning_Angel Agent and Semantic Agent that acts as supervisors constantly online to handle queries. Next, the mechanism also provides a QA subsystem that acts as assistant. The Learning_Angel can detect syntax errors written by the online users. The Semantic Agent can check the semantic of each sentence. Sometimes learners may make semantic level mistakes. This implies that they don't understand the course topic. The semantic agent can thus give some correction suggestions to users and analyze the data in the learner corpus. Moreover, when users query the system, the system will attempt to find the answer from the knowledge ontology or learner corpus. Besides, if sufficient number of QA pairs has been accumulated, the FAQ can act as a powerful learning tool for the learners.

Keywords: Semantic Agent, Link Grammar, and E-learning

1. Introduction

In recent years, Distance Learning has been undoubtedly the hottest issue in the discipline of computer science [1]. Aside from that, online learning technology through the web interface has been comprehensively developed in the decade. According the ability of incorporating with multimedia, WWW has become an ideal platform for distance learning [2]. With the rapidly technological pervasion in the Internet, Distance Learning allows students to enroll courses on the Internet and acquire new knowledge. It is a good solution for anyone who doesn't have enough time to attend traditional classes. Therefore, Distance Learning plays a very important role [3-7].

The advantage of Internet is information sharing. There are many kinds of applications over Internet to support information interchange, such as Telnet, FTP, E-mail, BBS and Chat-room. These applications need message exchange. Each participant can communicate with other participants through text-, voice-, and even video-based messages.

However, it is difficult for instructors trace the activities and behaviors of learners in distance learning environment. For examples, instructors may need to know:

- Do learners understand the context of teaching?
- Do learners talk about the indicated issues that instructors set up?
- Do learners understand the issues of course?

Therefore, it is quite useful if there are some automatic supervising mechanisms. This mechanism will monitor the dialogue and detect the mistakes on grammar and relevance of courses. It helps people obtaining completely education training without a centralize classroom. Thus, people can teach or learn at anywhere and any time.

However, there are many problems in distance learning systems such as: instructors cannot control the learners' activities, instructors cannot online forever, instructors don't keep the frequent answers and questions (FAQ) and thus learners cannot learn from ancestral learners and other learners.

To solve the problems mentioned above, the goal of this paper is to build up an ontology-based Semantic Agent system that provides supervising and learning-assistance for textual chat rooms. This system was built based on Agent, Link Grammar, XML, Learner corpus and other supporting functions to solve the instructor-off problems. The system provides Learning_Angel agent and Semantic agent. Also, the QA sub-system can collect/analyze the frequent mistakes and problems. The Learning_Angel agent was designed to provide the monitoring and syntax checking online. If learners fall behind the

Proceedings of the 25th IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW'05)
1545-0678/05 $20.00 © 2005 **IEEE**

Authorized licensed use limited to: Tamkang University. Downloaded on March 24,2010 at 02:03:46 EDT from IEEE Xplore. Restrictions apply.

discussing courses, Semantic agent can give some comments and/or suggestions. The statistical analyzer then records, classifies, analyzes the learners' dialogue. Furthermore, this dialogue can be used to generate QA pairs and update the learner corpus. Based on these corpora, instructors can revise or enhance their content of teaching materials. Learners can also learn from the experience of the ancestral learners and other learners.

This article is organized as follows: section 2 describes the related works that includes the introduction of Link Grammar and Ontology. Section 3 presents the architecture of proposed system. The chief processes of the proposed system will be shown in section 4. The last part of this article is the conclusions and future researches.

2. Backgrounds

2.1 Link grammar

Link grammar is an English grammar parser system that was proposed by the School of Computer Science of Carnegie Mellon University (CMU). Link grammar is a scheme to describing natural language [8]. Link grammar defines a set of words, which are the terminal of grammar, and each has some linking requirements. These linking requirements of each word are gathered in a dictionary. Figure 1 illustrates the linking requirements defined in simple dictionary for these words: a/the, cat/mouse, John, ran, chased.

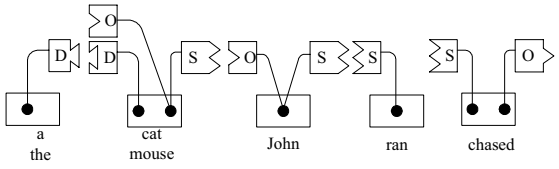

Figure. 1: words and connectors

Each intricately shaped box is a labeled connector. A pair of compatible connectors will join given that they correspond to the same type. For each black dot, only one connector can be selected. Figure 2 depicts how the linking requirements are satisfied in the sentence, "The cat chased a mouse."

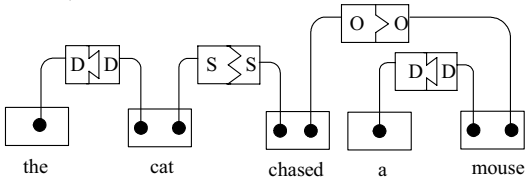

Fig. 2: All linking requirements are satisfied to form a linkage

The linking requirement for each word is expressed as a formula involving the operators &, and or, parentheses, and connector names. The + or – suffix on a connector indicates the direction in which the matching connector must be laid.

A sequence of words is a *sentence* of the language defined by the grammar. If links can be established among the words so as to satisfy the formula of each word, it forms a "linkage". The linkage includes the following meta-rules:

Planarity: The links are drawn above the sentence and do not cross.

Connectivity: The links suffice to connect all the words of the sequence together.

Ordering: When the connectors of a formula are traversed from left to right, the words to which they connect proceed from near to far. Namely, consider a word, and consider two links connecting that word to the word on its left. Comparing to the other word, the link connecting the closest word (the shorter link) must appear to the left (in the formula). The same process stands for a link to the right.

Exclusion: No two links may connect the same pair of words.

The use of a formula to specify a link grammar dictionary is convenient for creating natural language grammars. However it is cumbersome for mathematical analysis there of, and as well as in designing algorithms to parse link grammar. An alternate method of expressing link grammar is known as disjunctive form, which each word has an associated set of disjuncts. A disjunct is denoted as: **((L1, L2, …,Lm)(Rn, Rn-1, …, R1))**

Where L1, L2, …, Lm and Rn, Rn-1, …, R1 are the connectors that must connect to the left and right, respectively.

Thus, simply rewrite each disjunct and combine them with the 'or' operator to create an appropriate formula. Translating a link grammar from a disjunctive to a standard form can be accomplished as follows: **(L1&L2&…&Lm&R1& R2&…&Rn)**

Enumerating all ways that the formula cab be satisfied translates a formula into a set of disjuncts.

2.2 Ontology

Ontology is a popular research issues in various communities such as knowledge engineering, natural language processing, intelligent information integration and knowledge management. It provides a shared and common representation of a domain that can be communicated between heterogeneous and widespread application systems. Ontology has been

Proceedings of the 25th IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW'05)
1545-0678/05 $20.00 © 2005 **IEEE**
Authorized licensed use limited to: Tamkang University. Downloaded on March 24,2010 at 02:03:46 EDT from IEEE Xplore. Restrictions apply.

developed in AI to facilitate knowledge sharing and reuse. Ontology provides an explicit conceptualization that describes the semantics of the data. [12, 13]

Current computer systems are changing from single isolated device to entry points into a worldwide network of information exchange. Therefore, support in the data exchange, information and knowledge is becoming the key issue in communication technology. It can communicate between people and application systems. Providing shared and common domain structures is becoming essential, and used to describe the structure and semantic of information exchange. Now, Internet technology and the WWW are the main technology infrastructure for online information exchange. It's not surprising to see that a number of initiatives are arising in this area to provide notations for data structures and semantics. For example:

- The resource description framework (RDF)
- The Extendible Markup Language (XML)
- XML schemes provide a standard for describing the structure and semantics of data
- The transformation language of XSL (XSL-T)
- Various querying language for XML (XQL, XML-QL)

Depending on their generality level, different types of ontology may be identified that fulfill different roles. We can distinguish the following ontology types:

- Domain ontology: capture the knowledge valid for a particular type of domain.
- Metadata ontology: like Dublin Core [9] provides a vocabulary for describing the content of online information sources.
- Generic or common sense ontology: aim at capturing general knowledge about the world.
- Representational ontology: do not commit themselves to any particular domain.

3. System Architecture

This section introduces the theme of proposed Chat Room System, the diagram shows in figure 3. The left section of the diagram shows the components of Augmentative Chat Room, the flow of Chat Room Supervision and the Ontology Definition process. Then the right section is the database, which includes Distance Learner Ontology, Learner Corpus Database, and User Profile Database. Besides, the Questions and Answers System will analyze the Corpus and user profile to collect frequent questions that were made by learners. Finally the data will be sent to FAQ system for generating new OA-pairs.

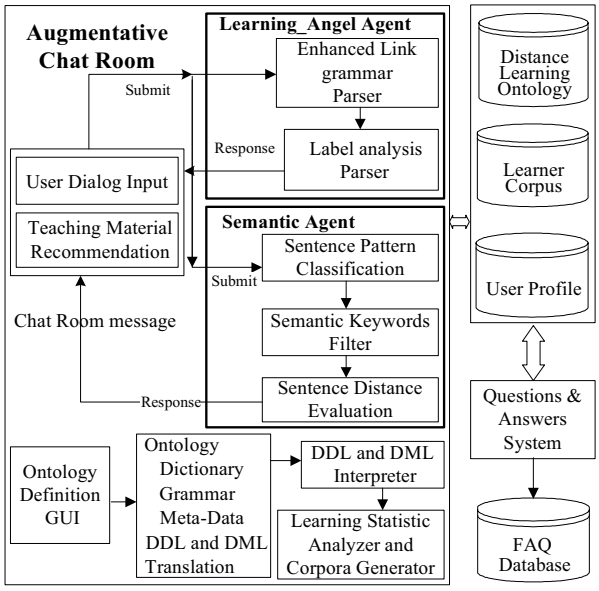

Fig. 3 The system architecture and operation flow

4. System Design

4.1 Domain Specific Sentences

Before introducing the each chief sub-system of the proposed chat room, we firstly describe the reason to restrict the research to specified domain. Domain specific sentences refer to those sentences that appear frequently in some certain application domain text but scarcely in others. There are some characteristics in domain specific sentences such as: 1) Vocabulary set is limited; 2) Word usage has patterns; 3) Semantic ambiguities are rare; 4) Terms and jargons appear frequently in the domain. [10]

It is fairly hard to apply semantic-level analysis to common language conversation. Take the following two sentences as examples, the syntax of these two sentences "The car is drinking water." and "The data is pushed in this heap." is correct. But the meaning of these sentences is incorrect. But in real worlds, the car can't drink water. In the data structure course, heap doesn't have push method. In fairy tale, cars maybe can drink water or others drink even cola. Therefore, in different situation the meaning of the sentence might be different. For this reason, the domain must be restricted. Thus the proposed system is deal with only "Data Structure" domain and same

Proceedings of the 25th IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW'05)
1545-0678/05 $20.00 © 2005 IEEE
Authorized licensed use limited to: Tamkang University. Downloaded on March 24,2010 at 02:03:46 EDT from IEEE Xplore. Restrictions apply.

scheme can be extended to other domain. By the way, the sentence that does not have syntax error but dose not make sense in the proposed chat room system is defined as "**Interrogative Sentence**".

According to the above reasons, the disusing topic and user messages are all restricted in the domain. Thus the terms of data structure are limited and can be pre-defined in the system ontology to support functions of syntax and semantic analysis.

Besides, system manager can load pre-define terms about Data Structure through the Ontology Definition GUI during system initialization. This Ontology Definition GUI interface is built to provide the ability that generates another scaffolding teaching material or extends the chapter for our courses. The ontology built in this system includes Dictionary, Grammar, and Meta-Data. This process of ontology creation is designed to translate the pre-defined ontology to DDL and DML form. At last, DDL and DML Interpreter can interpret ontology and then send the data to Corpora Generator, which records the data to Distance Learning Ontology and Learner Corpus databases.

4.2 Learning_Angel Agent

The following explains the Learning_Angel processes in figure 4. Strictly speaking, when the Augmentative Chat Room learners submit sentence to Learning_Angel Agent, Learning_Angel will forward it to Link Grammar Parser. Then the grammar parser will query ontology to get the tag about input sentence [11]. On the other hand, Link grammar parser will send the tags and sentence to Label analysis & filter, which can find if there have any mistaken linkages. In addition, if the input messages have some grammar errors, Label analysis & filter can also detect them and search for the suitable sentences from Learner Corpus and convey them to the online learners. By the way, Label analysis & filter will analyze the links of input words' sequences sent by Link Grammar Parser to check if the links of words satisfy the meta-rules such as planarity, connectivity, ordering and exclusion. If the input words' sequences have particular tag from Learning_Angel, the Label analysis & filter can record it in Learning Corpus, then efficiently send the correct information to the online learners.

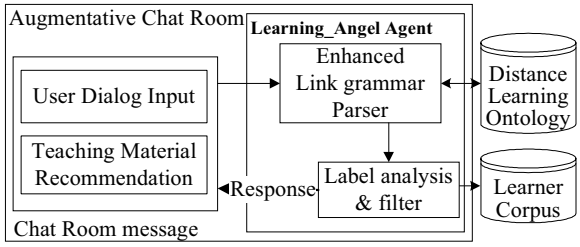

Fig. 4 Workflow of Learning_Angel Agent

4.3 Semantic Agent

This section describes the design of the Semantic Agent. Semantic Agent is designed to wisely find out the most appropriate explanations in learners' dialogues if learners said something that dose not make sense. There are two proposed methodologies of constructing Semantic Agent in this paper. One is Semantic Link Grammar that based on Link Grammar, and the other is Semantic Relation of Knowledge Ontology that is based on Ontology technology. Semantic Link Grammar can use the algorithm from link grammar to parse sentences. However, it is quiet difficult to modify the dictionary, which consists the correctly semantic meanings. It will take a lot of cost and time for linguistic classification and the performance is not very well.

In this paper, we had decided to choose the second methodology: Semantic Relation of Knowledge Ontology. This methodology is to evaluate the distance of specified keywords. The system flow follows three branching systems in turn:

1. **Sentence Pattern Classification**

   Firstly, the sentence pattern classification will check the sentence pattern. If there has an *Interrogative Sentence*, the syntactically incorrect, the system will ignore that sentence. Sometimes learners may not know the semantic meaning of this keyword, and they may need to ask other online users questions. Concerning this case, the system doesn't deal with the semantic problems. If there have no problems in a sentence, the sentence will be sent to next branching system.

2. **Semantic Keywords Filter**

   The system will detect five different kinds of sentences' patterns: 1) the Pattern in Simple Sentences, 2) the Pattern in Negative Sentences, 3) the Pattern in Question Sentences, 4) the Pattern in Sentences having WH questions, 5) the Pattern in Imperative Sentence. Then Semantic Keyword Filter will extract the sentence's keywords by using the information in Ontology.

3. **Sentence Distance Evaluation**

Authorized licensed use limited to: Tamkang University. Downloaded on March 24,2010 at 02:03:46 EDT from IEEE Xplore. Restrictions apply.

Semantic Keyword Filter will extract the sentence's keywords by using the Data Structure Ontology. For instance: ***The tree doesn't have pop method.*** Two keywords "tree" and "pop" will be recognized from the knowledge ontology. Then the system can search the words' *id*.

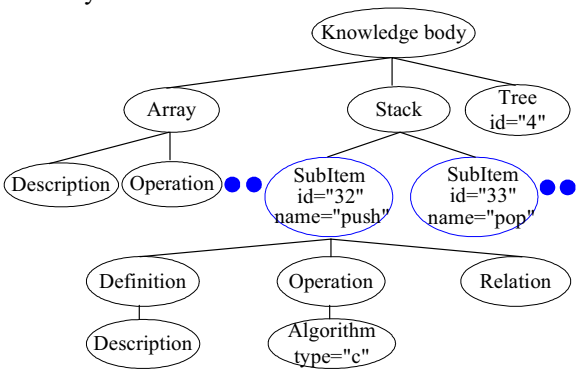

Fig. 5: The Data Structure Ontology Example

As figure 5 shows, the *id* of the keywords "tree" and "pop" is 4 and 33. We discover that the distance between two words is not related. But the pattern of this sentence is negative. The semantic checker will think this sentence is correct. The other possible *Interrogative Sentences* in the system might be:

- I push the data into a tree.
- The tree doesn't have pop method.

4.4 Questions and Answers System

In the last section, we decide to use the ontology-based approach- **Semantic Relation of Knowledge Ontology**- to construct semantic agent. This methodology can detect the sentence pattern and find the semantic-distance of the keywords in the knowledge ontology. According to these characteristics, that is a new ideal to design a Questions and Answers system (QA system). The workflow of QA system is shown in figure 6.

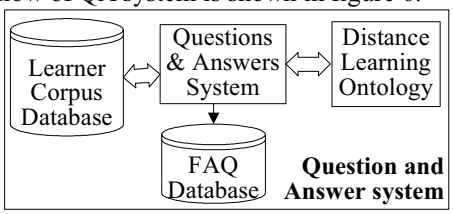

Fig. 6: Workflow of Questions and Answers system

When QA system gets an Interrogative Sentence pattern, the system will firstly can find the location of keywords in the ontology and then check semantic distance of the question by the semantic keywords and finally try to answer learner's question.

For examples:

What is Stack?

Which data structure has the method push?

Does stack have pop method?

When the QA system gets the sentence "**What is Stack**" from the learners, it will extract the keyword "stack" and find the positions of the knowledge ontology after matching the question template "What is". The system will know that the meaning of this question is to ask the definition of stack and then find the definition /description of stack to the user. Thus, the system will collect this question and answer into the FAQ database.

```
- <KeyItem id="3" name="stack">
  - <Definition>
      <Description>A stack is a Last In, First Out
          (LIFO) data structure in which all
          insertions and deletions are restricted to
          one end called a top. There are three basic
          stack operations: push, pop, and stack top.
          </Description>
      <Symbol name="top">A stack is a linear list in
          which all additions and deletions are
          restricted to one end which is called the top.
          </Symbol>
    </Definition>
```

There are some interrogative templates of the Question and Answer system such as: "What is", "The relations of", "Is … has …" and "Which … has". Moreover, FAQ database can also use the technologies of data mining to collect the question and answer pairs from the learner when they discuss in this system. If sufficient number of QA pairs has been accumulated, the FAQ system will make the statistic of the questions and answers and then gets the most frequency Question and Answer pairs. The system can also show these QA pairs to the learner. This can be a powerful learning assistant to online learners.

5. Conclusion

In this proposed system, learners can send messages to each other in English environment. They can discuss the course with each other and ask teacher questions. In this paper, we had designed the Learning_Angel Agent and Semantic Agent to be online supervisors. These two agents can automatically help learner to practice English conversation and discuss about the major course. The Learning_Angel Agent can automatically detect

Proceedings of the 25th IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW'05)
1545-0678/05 $20.00 © 2005 **IEEE**
Authorized licensed use limited to: Tamkang University. Downloaded on March 24,2010 at 02:03:46 EDT from IEEE Xplore. Restrictions apply.

syntax errors. Then the Semantic Agent will check the semantic of sentences in any dialogues if learners fall behind the discussing courses. Thus the online teachers and tutors are not always waiting students to submit their questions in any time. In a word, this can solve the instructor-off problems.

The Link Grammar, a word-based parsing mechanism, was designed to be an accurate grammar checker. But it fails to focus on the ability of fault tolerance. Differing from the Link Grammar, this system is particularly useful for non-native English speakers. Besides, the proposed system can also parse sentences and give some comments or corrections to users. However, the original Link Grammar does not have this function and it appears that the proposed idea herein can be applied in other domain specific application.

With servicing the system constantly online, it is important for philologists to analyze sentences accumulated from students' dialogues. Then the system can easily point out the common or specific mistakes. Subsequently, online teachers can refine their learning materials.

In a conclusion, this system provides a better and more interactive environment for teachers and students. Words are made of a basic communicating unit that is common in natural language texts and speech-processing activities. When teaching English, teachers always want to know the route of mistakes that students may make. Thus, according to various functional necessities, the proposed system can be extended to encompass more scalable ranges. The notions presented herein can aid in the development of other similar application. In the future, we will focus on finding the better solutions to semantic analysis, evaluating the accuracy of the proposed Semantic Agent and finally trying to follow some famous distance-learning standards.

Reference:

[1] Tsang, H.W.; Hung, L.M.; Ng, S.C. "A multimedia distance learning system on the Internet", proceedings of IEEE International Conference on Systems, Man, and Cybernetics (SMC), Vol. 2, Page(s): 243 –246

[2] Adhvaryu, S., and Balbin, I., "How Useful is Multimedia on the WWW for Enhancing Teaching and Learning?", Proceedings of International Conference on Multimedia Computing and Systems (ICMCS'98), Austin, USA, June 28-July 31, 1998

[3] Harris, D., Cordero, C., and Hsieh, J., "High-Speed Network for Delivery of Education-on-Demand", Proceeding of Multimedia Computing and Networking Conference (MCN'96), San Jose, California, January 29-31, 1996

[4] Willis, Barry, "Distance Education at a Glance", Engineering Outreach at the University of Idaho, http://www.uidaho.edu/evo/distglan.html

[5] Goldberg, M. W., and Salari, S., "An Update on WebCT (World-Wide-Web Course Tools) – a Tool for the Creation of Sophisticated Web-Based Learning Environments", proceedings of NAUWeb'97 – Current Practices in Web-Based Course Development, Flagstaff, Arizona, June 12-15, 1997

[6] Goldberg, M. W., Salari, S., and Swoboda, P., "World Wide Web Course Tool: An Environment for Building WWW-Based Courses", Computer Networks and ISDN Systems, 28 (1996)

[7] Goldberg, M. W., "Student Participation and Progress Tracking for Web-based Courses Using WebCT", proceedings of the Second International N.A. Web conference, Fredericton, NB, Canada, October 5-8, 1996

[8] D. K. Sleator, Davy Temperley "Parsing English with a Link Grammar" October 1991, CMU-CS-91-196

[9] S. Weibel, J. Gridby, and E. Miler: OCLC/NCSA Metadata Workshop Report, Dublin, EUA, 1995. http://www.oclc.org:5046/oclc/research/conferences /metadata/dublin_core_report.html

[10] Jianming Li, Lei Zhang and Yong Yu, In the Workshop on Knowledge Markup and Semantic Annotation at the 1st International Conference on Knowledge Capture (K-CAP 2001), October, 2001, Victoria, B.C., Canada

[11] Ying-Hong Wang, Chin-Hao Lin,"A multimedia database supports English distance learning", An International Journal Information Sciences, Vol.158, Jan. 2004, Pages 189-208

[12] Deborah L. McGuinness and Frank van Harmelen, "OWL Web Ontology Language Overview" W3C Recommendation 10 February 2004

[13] Labrou, Y. and Finin, T, "Yahoo! as an ontology—Using Yahoo! Categories to describe documents," Proceedings of the 8th International Conference on Information and Knowledge Management. Kansas City, MO, Nov. 1999, pp. 180–187.

Authorized licensed use limited to: Tamkang University. Downloaded on March 24,2010 at 02:03:46 EDT from IEEE Xplore. Restrictions apply.