A Question-Answering Environment for eLearning Tutors

Joaquim Moré
Open University of Catalonia (UOC)
Av. Tibidabo 39-15. 080350 Barcelona. Spain

jmore@uoc.edu

Salvador Climent
Open University of Catalonia (UOC)
Av. Tibidabo 39-15. 080350 Barcelona. Spain

scliment@uoc.edu

Marta Coll-Florit
Open University of Catalonia (UOC)
Av. Tibidabo 39-15. 080350 Barcelona. Spain

mcollfl@uoc.edu

José Manuel Rivera
Open University of Catalonia (UOC)
Av. Tibidabo 39-15. 080350 Barcelona. Spain

jrivera@uoc.edu

Abstract— **In this paper we present a tool, which is being developed at the Learning Technologies Department of the Open University of Catalonia (UOC), to help tutors to answer questions e-mailed by their students in eLearning environments. The tool displays multilingual contexts taken from course materials, forums, Wikipedia and scholar papers in the Internet that may help the tutor to answer the students properly. The system allows tutors to find better answers as well as to update their knowledge and value the contribution of students in life-long-learning**

Keywords— **e-Learning, Question-Answering Systems, Speech Acts, Tutor**

I. INTRODUCTION

The UOC (www.uoc.edu) is a leading Spanish virtual University currently offering 33 university degrees, 10 Master's degrees 1 Ph.D. program, and several dozens of other courses. Communication between students, professors and tutors is completely performed via email or forums within a virtual campus and a system of virtual classrooms. Courses are taught in Spanish and/or Catalan.

Tutors are generally overwhelmed with students' questions and are unable to give a timely response. Question-Answering systems (QAS) have been developed in order to help tutors to answer properly and promptly. These systems generally reply the question automatically ([1], [2], [3]) but there are difficult aspects for a human-being emulator. One of these aspects is question-identification in a message where the question is not declared explicitly but implicitly, with deviations from formal and normative expression. Another problem concerns

questions about matters that are related to the academic subject but they are not explained in the course materials. In this case, the answers cannot be straightforwardly retrieved from Question/Answer-Pair databases and course- dependent ontologies ([2], [3], [4]). A way to overcome this drawback is to retrieve answers from Social Media Content ([5]) but, having into account the possible weird answers that can be found there, the system should learn to discriminate between good and bad answers, which is still too highly demanding for a state-of-the-art QAS.

Apart from these problems, which make tutors feel reluctant to trust automatic QAS, these systems do not take into account a process we have noticed in UOC's tutors' messaging. Many times the questions trigger the tutors to search for information and learn new things, thus updating their knowledge. Many questions arise from the student's reflections on exercises or course materials, so the tutor may be challenged to find a proper answer for an aspect she is not fully confident in, she has not considered before or, plainly, she did not know.

In this paper we present a semiautomatic tutor-assistant whose aim is not to retrieve an exact answer for a question, but provide the tutor with useful information to answer the student properly and promptly. The system also allows teachers to update their knowledge, and helps them to value the pupils' contribution in their life-long learning.

The paper is organized as follows. In the second section, we present the methodology based on a pragmatic theory. In the third and fourth sections we explain the prototype we have developed so far, and discuss its evaluation. Finally, we show the conclusions and the future work.

II. METHODOLOGY

The target user of the system consists of the tutors of the virtual classrooms of all UOC's degrees and programs. Therefore we have developed a scalable methodology which is independent of particular subject domains. We decided to tackle the problem from a theoretical framework, concretely the speech-act theory ([6]), which describes the basics of communication between an addresser and an addressee.

In a communicative situation such as the tutor-student's mailing, the student has a goal which is expected to be accomplished by the tutor. So it is crucial for the student to perform a speech act whose linguistic features make her expectation clear for the tutor. Conversely, the tutor's speech act embodied in the reply has to contain linguistic features that confirm the student that she is fulfilling the expectation.

In a speech act, there are two elements. The first one is the speech act expression (SAE); that is, the expression by which the addressee identifies the addresser's expectation. *I don't understand* is an example of how the student expresses her expectation to be clarified. The second element is the speech act objects (SAO), which is the term, or terms, focused by the speech act of the addresser. For example, if the student says *I don't understand the notion of c-command, c-command* is the SAO.

We hypothesize that document segments that are useful for tutors contain the SAOs of the message, and we wondered to what extent this SAO-sharing between message and document segments contributed to the usefulness of the system. Hence, the prototype we have developed searches for contexts in reliable information sources where SAOs (co-)occur. SAO candidates are automatically detected. However, the tutor selects the most relevant ones since she is able to grasp the real intention of the student despite the fuzzy discourse relationships that characterize informal e-mails.

The speech-act approach has also been taken in a work such as [1], but it depends on messages posted in other discussions, manually annotated with part-of-speech tags. However, for operative reasons, we could not consider annotating the messages that all UOC tutors have received for years.

III. THE PROTOTYPE

In this section we will explain the workflow of the prototype and the information sources it consults.

The workflow of the system consists of four stages:

1. Extraction of the subject of the message

The subject of the message is extracted in order to retrieve messages from discussions with the same or similar subjects.

2. Parsing of the message body

The system parses the body with the linguistic parser FreeLing [7]. As most of the messages are written in Catalan, this language is the default source language.

3. Tag cloud generation

The system presents a tag cloud in order for the user to select the SAOs. The tag cloud is generated by an automatic term extractor. Verbs, nouns, named entities, and quoted expressions are extracted from the tagged message. Quoted expressions are extracted because we consider that quotes are a typographical feature that the writer uses to make clear what his/her expectation is. Terms that are likely to belong to the subject domain are highlighted. However these terms are identified by using a subject-independent method. The system consults the downloadable, open-source Catalan-English DACCO dictionary (http://sourceforge.net/projects/dacco/), which has information about the frequency of the entries, according to their Google number of results. Hypothesizing that subject-domain terms have less results than general vocabulary terms, the terms highlighted are those below a threshold number of results.

4. Useful contexts search

After ticking the speech act objects in the tag cloud, the system searches for and displays contexts in Catalan and English where the denominations of the objects ticked (co)appear in these two languages. We will refer to these contexts as *useful context candidates* (UCC)

The information sources consulted to retrieve UCCs are the following:

1. Messages in forums posted in previous discussions

The question may have been formulated in previous discussions, and a student or another tutor may have posted a good answer. Unfortunately, all the messages posted during UOC's history are not stored. Only the messages received in the last four years were available when this project was initiated.

2. UOC's course materials

The system queries a search-engine developed at the UOC in order to find contexts in UOC's course materials where the terms selected by the tutor co-appear.

3. Wikipedia

Wikipedia entries in Catalan and English where the terms selected by the user are explained.

4. Online academic papers

The system queries the Delicious' search-engine in order to find papers in Catalan and English whose tags intersect with those selected by the tutor. The system also queries the CiteULike search engine (http://www.citeulike.org/), a free online service that organizes academic publications, and retrieves articles by the same method. So the system displays Delicious and CiteULike results pages with links to articles that deal with the concepts selected.

5. Internet pages

Internet pages retrieved by the Yahoo Search engine, where all the terms selected by the user co-appear.

III. EVALUATION

The prototype was evaluated for the academic subject *Linguistics II*, and the evaluators were 6 people, divided into two groups. The members of the first group were experienced tutors of the subject. We will refer to this group as the *experienced group*. The second group, the *novice group*, were graduate and teachers in linguistics that had not had any experience as tutors before. We wanted to compare the impression of the system of experienced users and the impression of novices, in order to know how the system can help novices to perform a task that is new for them.

40 messages were selected for the evaluation. The number of messages and the semesters they covered- the two latest- were due to the fact that, on the one hand, they could not be stored in the previous discussions database. On the other hand, we dismissed those messages that at least 3 evaluators agreed in considering so decontextualised and underspecified that no useful context could be retrieved for them.

For each message, the prototype displayed the UCC found in each information source according to the terms selected in the tag cloud. Both the experienced and the novice group scored the usefulness of the information source according to the UCC displayed.

The usefulness of the system had two dimensions. The first one was the usefulness of the information source to provide a good answer (UPGA). The second dimension was the usefulness of a helpful information source to provide a prompt answer (UPPA). Both items were scored with a five-value scale: 0 (not useful), 1 (not very useful), 2 (useful), 3 (very useful), and NC (no context), meaning that the system could not retrieve any UCC from the information source. A good UPGA score was implied in information sources that had a good score for the usefulness to provide a prompt answer (UPPA).

The evaluators were encouraged to write comments about the efforts and difficulties in getting helpful contexts. These comments provided us with useful information to improve the system.

The evaluation analysis was split into a *macroevaluation* and a *microevaluation*. The macroevaluation consisted in scoring, for both the expert and the novice group, the usefulness of the system in providing a good answer and in providing a prompt answer. The microevaluation was intended to draw information about the contribution of each information source in the usefulness of the system, and also to find aspects of the system to be improved.

A. Macroevaluation

Although the contribution of each source of information was evaluated separately, if one source had the highest score as useful to provide a good answer, the system was considered useful for the evaluator because she found what she was looking for.

The usefulness to provide a good answer score (UPGA), according to the expert group, was calculated as follows: we collected the scores of an expert and, for each message, we took the score of the most valued source of information (top score). Then, we calculated the evaluator's mean score (EMS), which was the average of the top scores. The UPGA was the mean of the EMS of the three evaluators.

We calculated the UPGA, according to the novice group, by using the same method. Besides, the UPPA, according to the experienced and the novice group, was calculated in the same way.

In a range from 0 to 3, the experienced group scored higher the usefulness to provide a good answer (1.733) than the novice did (1.517). On the other hand, the novice group scored higher the utility to provide a prompt answer (1.75) than the experienced group (1.467).

B. Microevaluation

For the experienced group, snippets from web pages contributed the most to provide a good answer, with a considerable distance from the other sources of information. For the novice group, snippets from web pages and Wikipedia articles were on the top list, and the distance from the other sources was closer.

According to the comments made by expert tutors, although they found useful contexts from web pages, their discrimination took them a lot of time. On the other hand, they said that they found useful contexts after more than one try by selecting different terms from the tag cloud. This explains, somehow, the worse score for the utility to provide a prompt answer we have seen in the expert tutors.

Course materials are ranked as the second useful source of information for experienced tutors. So it seems that the ranking difference from the novices lies in the experience of the experts in searching for information related to the subject by using search engines, and their ability to combine keywords to get useful results.

Papers were on the bottom of the list, below the 1.5 score for both the expert and the novice group. According to a comment of the evaluators, this may be due to the fact that papers find in Delicious and CiteULike deal with very specialized topics and their target consists mainly of professors and post-graduate students. On the contrary, the topics dealt with in Wikipedia fit better the questions and reflections of under-graduate students.

A fact we have detected is the mapping between types of students' expectations and information sources. For example, course materials were useful to clarify a concept, although they were not always the highest ranked because students generally ask for information not found in the materials. On the other hand, Wikipedia's references and links to external online resources were useful to find a solution to a problem, or suggest further reading. Besides Wikipedia articles and web pages provided extra information that complemented the information of the learning material, and helped the student to confirm that her reflections, even digressions, were on the right track. So these sources of information are useful for the tutor to update her knowledge. Even scholar papers, if they dealt with the topics of the course material, would contribute to the tutor's life-long learning.

Previous messages posted in forums fit messages where the student asks for help. However, their usefulness depends on how recurrent the problem has been along the course's history.

On the other hand, previous messages cannot be retrieved if the student writes with underspecified references that are only meaningful in the temporal context they were written.

IV. CONCLUSIONS AND FUTURE WORK

In this article we have presented a tutor assistant whose methodology is flexible enough to tackle other student's communicative goals apart from answering a question. This is a distinctive feature from traditional approaches. When messages contain digressions or reflections evoked by the reading of the materials, the system fosters a 'learn from your pupils' learning process. A process which is evident when the tutor finds that a student posted a good answer in a previous forum.

The results of the prototype evaluation are quite promising, regarding the short period of time spanned in the evaluation corpus, and the dependence on how recurrent questions, reflections and digressions are. However, the methodology we have developed so far involves spending time and slightly favors experienced tutors. So our goal is to improve the usefulness to get a prompt answer and improve the usefulness to provide a good answer for experts and novices alike.

We are thinking of improving the usefulness to provide a prompt answer by widening the pragmatic approach, and take advantage of the mappings between the student's expectations and the information source. Moreover, the mappings between the speech-act expression of the student and the speech-act expression in the text segment that best fits her expectation will be taken into account.

On the other hand, we are planning to ease the searching for useful contexts for experts and novices alike by keyword expansion. That is, the terms selected by the user will trigger terms that are semantically close-related to them, although they are not visible in the tag cloud.

Last but not least, we are thinking of integrating a search engine for scholar documents with useful contents for graduate students and tutors alike. PhD dissertations, especially their state-of-the-art sections, are interesting candidates.

ACKNOWLEDGMENTS

REFERENCES

[1] D. Feng, E. Shaw, J. Kim and E. Hovy, "An Intelligent Discussion-bot for Answering Student queries in threaded discussions", in *Proceedings of Intelligent User interfaces.* 2006.

[2] C.Y. Yang, "A Semantic FAQ System for Online Community". *Journal of Software.*, vol. 4, num 2. 2009. pp. 153–158.

[3] J.C. Hung, C.-S. Wang, C.-Y. Yang, M-S. Chiu, and G. Yee, "Applying Word Sense Disambiguation to Question-Answering System for e-Learning." In *Proceedings of the IEEE 19th International Conference on Advanced Information Networking and Applications*. 2005. Taiwan. pp. 157-162.

[4] Y. Wang, W. Wang, and C.Huang, "Enhanced semantic question answering system for e-learning environment." In *Proceedings of the 21st international conference on Advanced Information Networking and Applications Workshops*. 2007(2), pp. 1023-1028.

[5] D. Bernhard and I. Gurevych, "Answering Learner's Question Paraphrases from Social Q&A Sites" In *Proceedings of the 3rd Workshop on Innovative Use of NLP for Building Educational Applications, ACL.* Columbus, Ohio. 2008. pp. 44- 52

[6] J. Searle, *Speech Acts. An Essay in the Philosophy of Language*. Cambridge, 1969

[7] J. Atserias, B. Casas, E. Comelles, M. González,   L. Padró, M. Padró, "FreeLing 1.3: Syntactic and semantic services in an open-source NLP library "In *Proceedings of the fifth international conference on Language Resources and Evaluation (LREC 2006)*, ELRA. Genoa, Italy. May, 2006.