ABOUT THE BENEFITS OF EXPLOITING NATURAL LANGUAGE PROCESSING TECHNIQUES FOR E-LEARNING

Diana Pérez-Marín, Ismael Pascual-Nieto and Pilar Rodríguez
Computer Science Department, Universidad Autonoma de Madrid
Francisco Tomas y Valiente, 11, 28049, Madrid, Spain

Keywords: Natural Language Processing, Adaptive Hypermedia, User Modeling, Open Learner Modeling, Conceptual Modeling, e-assessment, e-learning.

Abstract: Natural Language Processing (NLP) is a research field that studies how to automatically interpret/generate information in natural language. Currently, the quality and number of developed NLP resources and techniques permit their application to educational systems, with the potential of widening access to training and opening new ways of teaching. In this paper, the benefits of exploiting the current NLP techniques to improve e-learning systems will be discussed. A brief overview of the state-of-the-art of NLP will be provided and, some real e-learning and e-assessment applications based on the use of NLP techniques will be described to illustrate the benefits of using NLP techniques for e-learning.

1 INTRODUCTION

In the past there were only three possible ways to acquire new information: by attending to lessons in the school, high-school or university; by extracurricular activities; or, by self-teaching with books or manuals.

Since the 80s, with the Multimedia Age this fact started to change and the creation of CD-ROMs with static information became a new approach to learning. These courses on CD-ROM were simply the textbooks typed into the computer and there were problems because once the data was written on the CD-ROM it could not be altered and therefore updating the course meant delivering new CD-ROMs with the high cost it implied.

Experts became aware of the possibilities Internet had as a teaching tool. On-line education was invented and e-learning started to be used by more and more teachers and students.

E-learning systems allow students to control their own learning rhythm. They can connect from anywhere at anytime. However, the lack of a tutor can become a problem if the student gets stuck in a lesson.

In order to approach this and to provide the benefits of one-on-one instruction (pay attention to each student's learning needs, assess and diagnose problems and provide assistance as needed), Intelligent Tutoring Systems (ITS) were created.

ITSs are computer-based teaching or training systems which reduce cost by automating course work selection, presentation, and student evaluation. One key factor of any ITS is to keep a student model with information from the student to tailor strategies and provide explanations, hints, examples, and practice as needed.

Inspired by ITSs, in the early 1990s, Adaptive Hypermedia Systems (AHS) were born. They combine hypermedia-based systems with adaptive and user-model-based interfaces (Eklund and Sinclair, 2000).

They can be used in any situation in which several users with different learning styles and backgrounds have to access common information. The adaptation should be not only static (based on stereotypes) but also dynamic (based on the student's behavior). It has proven to be effective since learners using such systems have demonstrated faster learning, more goal-oriented attitude and take fewer steps to complete a course (Conlan, 2003).

There are several educational AHSs (AEHSs) that are currently being used both in academic and commercial environments (Brusilovsky, 2004). However, all of them rely on objective testing items to assess the students' knowledge.

Pérez-Marín D., Pascual-Nieto I. and Rodríguez P. (2008).
ABOUT THE BENEFITS OF EXPLOITING NATURAL LANGUAGE PROCESSING TECHNIQUES FOR E-LEARNING.
In *Proceedings of the Fourth International Conference on Web Information Systems and Technologies*, pages 472-475
DOI: 10.5220/0001532304720475

On the other hand, Natural Language Processing (NLP) has studied since the 50s how to automatically analyze, extract information and generate natural language. Although the advances in the field have not been as spectacular as they were expected at the beginning, the current number of NLP techniques and resources permit their application to e-learning systems. For instance, AEHSs could take advantage of Information Extraction techniques to improve the student model and thus, the adaptation to his or her particular features; or, the evaluation section of e-learning courses could incorporate open-ended questions by using free-text Computer Assisted Assessment techniques.

In this paper, the benefits of exploiting the current NLP techniques and resources to improve e-learning systems will be discussed. In order to do that, the paper is organized in four sections: Section 2 gives a brief overview of the state-of-the-art of NLP: Section 3 describes some real e-learning applications that rely on some kind of NLP technique or resource; and, finally Section 4 ends with the conclusions.

2 OVERVIEW OF THE STATE-OF-THE-ART OF NLP

Linguistics is the field that studies language. Computational linguistic researches linguistics phenomena that occur in digital data. Natural Language Processing (NLP) is a subfield of Computational Linguistics that focuses on how to build automatic systems able to interpret/generate information in natural language (Volk, 2004).

A brief historical review would start in the fifties. It is important to highlight Chomsky as the provider of the theoretical basis for the following systems. In general, it was a time of big expectations in which several classes of grammar were characterized, the Probability theory of Shannon was formalized and the first NLP application (Machine Translation) was explored. However, the systems that were created were very simple. They just translated word by word accessing dictionaries. The task became more complex than originally thought. This culminated with the ALPAC (Automatic Language Processing Advisory Committee) report at the end of the sixties that stated that the automatic translation of scientific texts did not exist and that it was not expected soon. This implied reducing the funding in NLP not only in USA but also in other countries.

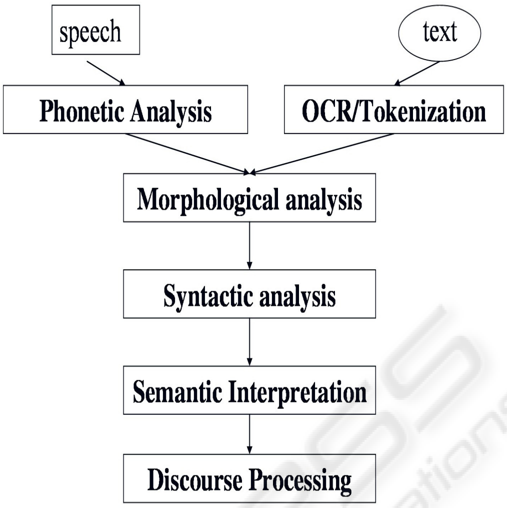

Figure 1: Typical NLP pipeline.

The following decades were characterized by the realism. In fact, two important ways of processing were treated: symbolic and statistic. The former relies on methods of qualitative analysis such as hand-crafted rules and the latter uses the distribution of quantitative text features to draw conclusions.

To build the rules of analysis is complex and takes time. On the other hand, in order to achieve a good performance, the statistical approach needs of a big amount of annotated texts (corpora). Some researchers advocate the possibility of having hybrid systems relying on a combination of statistical and symbolic techniques (Hermet and Szpakowicz, 2006).

A typical NLP system is represented in Figure 1. It can be seen how in order to process natural language, several linguistics levels should be covered:

- Phonological: Sounds processing to detect expression units in speech.
- Morphological: Extracting information about the words such as their gender, number and part-of-speech. Besides, considering suffixes, prefixes and other derivational, inflectional and compositional issues.
- Syntactical: Using parsers to detect valid structures in the sentences. These structures are usually represented with graphs or trees. It serves as basis for semantic interpretation.
- Semantic: Finding the most suitable knowledge formalism to represent the meaning of the text.
- Pragmatic: Interpreting the meaning of the sentence in a context to react accordingly.

Many resources have been developed in order to be able to cover as many as possible levels. Some of them are electronic dictionaries, thesauri, ontologies (one of the most relevant is WordNet for English and EuroWordNet for European languages), morphology and spelling rules, grammar rules, semantic interpretation rules and discourse interpretation templates.

Some NLP tools are tokenizers to break the text in tokens (a unit of processing), part-of-speech taggers, chunkers, sentence splitters, parsers and logic translators. They can be applied to solve different problems such as:

- Machine Translation, e.g. Babelfish (Alta Vista).
- Question Answering, e.g. Ask Jeeves (Ask).
- Language Summarization, e.g. MEAD (U.Michigan) .
- Automatic Essay evaluation, e.g. E-Rater (ETS).

The main challenges are now in the semantics and pragmatics levels. For them it is still necessary more research on general versus domain-specific resources and algorithms, the interplay between prosody, syntax, and semantics, new means of communication and new types of discourse.

3 APPLICATIONS

In this section, some applications will be described to illustrate the potential of using NLP techniques for e-learning systems. The goal is to provide a representative list, highlighting several advantages of using NLP techniques for different applications:

- CarmelTC (Rosé et al. 2003): Carmel is a Virtual Learning Environment system that has been incorporated the free-text assessment module called CarmelTC. It is able to give a score to the student and, to find out which set of correct features are present in student essays by using the Carmel's linguistic analysis of the text. The advantage of using NLP techniques is to allow the assessment of higher cognitive skills than just limiting the evaluation to objective testing such as Multiple Choice Questions or fill-in-the-blank exercises.

- Welkin (Alfonseca, 2003): It is a web-based application based on the wraetlic NLP toolkit to automatically select the contents of a text according to a user profile. The advantage of using NLP in Welkin is to allow the user to focus just on the snippets of texts that are interesting for him or her. For instance, if the profile of the user indicates that s/he is only interested in historical events, information about non-historical events will not be shown to him or her.

- E-tester (Guetl et al., 2005): It is a computer-based system that identifies the main concepts in a text and generates questions from these concepts such as "What is xxx?" or "Explain yyy". Next, it waits for the students' answers in free-text to compare them with the e-learning content that the system has and treats as model answer. The comparison is based on the free-text scoring system Markit (Williams and Dreher, 2005). The advantage of using NLP is to avoid the necessity of asking the teacher to introduce the questions as they are generated by the system. Besides, the students are able to see a histogram of frequencies indicating how well they have used each concept in comparison to the number of times the concept appears in the model answers.

- AutoTutor (Graesser et al. 2007): It is a web-based intelligent tutoring system able to engage the student in a dialogue by using NLP techniques. It appears as an animated agent that acts as a dialog partner with the student. The advantage is that the student is more motivated to continue studying the subject and find the task not only interesting but amusing as s/he feels that s/he is talking to a person.

- Computing similarity between users (Lops et al. 2007): It is a system that clusters users' profiles using WordNet. It has the advantage of making the comparison and clustering of good and bad aspects of the users' profiles easier.

- Knowledge Tracing (Corbett et al. 2007): Traditionally, Intelligent Tutoring Systems have incorporated student explanations of problem solutions in menus. The advantage of using knowledge tracing of typed student explanations is that the model created is able to predict better the student performance.

- Question answering (Heiner, 2007): It is a system that is being developed within a Ph.D. thesis to allow students to ask questions to the e-learning system. The advantage of using NLP is that the system is able to classify the questions and try to give an adequate answer.

- TAGARELA (Amaral and Meurers, 2007): It is a Computer Assisted Language Learning (CALL) system, which is being developed within the project with the same name, and the advantage is to be able to provide individualized language instruction.

- Will Tools (Pérez-Marín, 2007): It is a set of web-based applications that are able to automatically generate a student's conceptual model from his or her answers in free-text. The Will Tools consist of: Willed, the authoring tool to create the questions to ask the students (each question has a statement, maximum score and several correct answers written by the teachers); Willow, the system to ask the questions introduced in Willed to the students (the core idea is to compare the student's answer to the teachers' answers and the more similar they are, the bigger the score provided to the student); Willov, the conceptual model viewer (a conceptual model can be defined as a network of concepts that can be visually displayed as a concept map, conceptual diagram, table, graph or textual summary); and, Willoc, the configuration tool. The advantage of using NLP in the Will Tools is not only to automatically score free-text students' answers, but also to permit the estimation of a confidence-value for each concept used by the student in his or her answer and thus, to provide feedback to students and teachers about how well each concept is known.

4 CONCLUSIONS

In this paper, it has been claimed that e-learning systems should take advantage of the currently available NLP techniques and resources. In particular, there have been reviewed several educational applications in which NLP techniques have been applied and, in each of them, the advantage of using NLP to improve its functionality has been highlighted. More applications of NLP to e-learning can still be explored to open new ways to distance education.

ACKNOWLEDGEMENTS

This work has been sponsored by Spanish Ministry of Science and Technology, project TIN2007-64718.

REFERENCES

Alfonseca, E. 2003. *An Approach for Automatic Generation of on-line Information Systems based on the Integration of Natural Language Processing and Adaptive Hypermedia techniques*, PhD thesis, Computer Science Department, Universidad Autónoma de Madrid, Spain.

Altavista, http://babelfish.altavista.com/.

Amaral, L. and Meurers, D. 2007. Designing Learner Models for Intelligent Language Tutors. *CALICO 2007* May 22-26, Texas, U.S.A.

Ask, http://www.ask.com/.

Brusilovsky, P. 2004. Adaptive educational hypermedia: From generation to generation, in *Proceedings of the 4th Hellenic Conference on Information and Communication Technologies in Education*.

Conlan, O. 2003. *State of the art: Adaptive hypermedia, Technical report*, Waterford Institute of Technology.

Corbett, A., Wagner, A., Lesgold, S., Ulrich, H. and Stevens, S. 2007. Modeling Students' Natural Language Explanations, in *Proceedings of the User Modeling international conference*, Greece.

Eklund, J. and Sinclair, K. 2000. *An empirical appraisal of adaptive interfaces for instructional systems*, Educational Technology and Society Journal, 3.

ETS, http://www.ets.org/

Graesser, A.C., Jackson, G.T., and McDaniel, B. 2007. *AutoTutor holds conversations with learners that are responsive to their cognitive and emotional states,* Educational Technology, 47, 19-22.

Graham, C. 2006. *The Handbook of Blended Learning: Global Perspectives, Local Designs*, Pfeiffer, chapter Blended Learning Systems.

Guetl, C., Dreher, H. and Williams, R. 2005. *E-tester: A computer-based tool for auto-generated question and answer assessment*, E-Learn journal.

Heiner, C. 2007. Towards A Virtual Teaching Assistant to Answer Questions Asked by Students in Introductory Computer Science, in *Proceedings of the Artificial Intelligence for Education international conference.*

Hermet, M. and Szpakowicz, S. 2006. Symbolic assessment of free text answers in a second-language tutoring system, in *Proceedings of the 10th Computer Assisted Assessment conference*, Loughborough, United Kingdom.

Lops, P., Degemmis, M. and Semeraro, G. 2007. Improving social filtering techniques through WordNet-based user profiles, in *Proceedings of the User Modeling international conference*, Greece.

Pérez-Marín, D. 2007. *Adaptive Computer Assisted Assessment of free-text students' answers: an approach to automatically generate students' conceptual models*, PhD thesis, Computer Science Department, Universidad Autónoma de Madrid, Spain.

Rosé, C.P., Roque, A., Bhembe, D. and VanLehn, K. 2003. A hybrid text classification approach for analysis of student essays. In HLT-NAACL Workshop on Building Educational Applications Using Natural Language Processing, pages 68–75.

Shute, V. and Torreano, L. A. 2002. Formative evaluation of an automated knowledge elicitation and organization tool, *Authoring Tools for Advanced Technology Learning Environments: Toward Cost-Effective Adaptive, Interactive, and Intelligent Educational Software* .

U. Michigan, http://www.summarization.com/mead/

Volk, M. 2004. *Introduction to natural language processing*. Course CMSC 723 / LING 645 in the Stockholm University, Sweden.

Williams, R. and Dreher, H. 2005. Formative assessment visual feedback in computer graded essays, in *Proceedings of the Issues in Informing Science and Information Technology* (INSITE), Arizona, U.S.A.