

Faculty Research Grant – Summer 2004

An Evaluation System for Logic Form Identification

Vasile Rus

This grant, first of all, helped with making IUSB viewed as a university dedicated to high quality and high-end research on a global scale. People from famous universities and companies from all over the world were involved as active participants in some of the activities sponsored by this work and used the resources made available by this work. The increased visibility resulted from this project for the Computer and Information Sciences department will also have long lasting benefits.

The proposal developed a package for automatic evaluation of intelligent systems that use logic form as their underlying logic representation. The general evaluation method was to develop a *gold standard*, a set of test data annotated with the correct logic form, and then automatically compare the gold standard with the output given by a certain intelligent system. The **outcomes** of the project are: a paper published at ACL 2004 (Annual Meeting of Association for Computational Linguistics Conference) in Barcelona, Spain, a workshop organized with participants from all over the world from academia and industry, a first software package for automatic evaluation of Logic Form Identification methods, a first gold standard that is and will be widely used and referenced throughout the community, a journal paper submitted to International Journal of Computers and Applications.

The **objectives** of the proposal were successfully achieved:

- Test and trial data of about 500 sentences were carefully selected to be representative for different linguistic of English. The data sets were then annotated with correct logic forms and few quality assurance steps were applied to deliver high-precision of the annotation. The annotated version is *the gold standard* against which outputs of different software systems were compared. The test and trial data is available here:
<http://www.cs.iusb.edu/~vasile/logic/indexLF.html>
- Two new metrics, Precision and Recall, were proposed and details about them can be found on this web page:
<http://www.cs.iusb.edu/~vasile/logic/indexLF.html>
- A software package that automatically scores different systems was developed and made publicly available for use by the entire community. The software is available for download here:
<http://www.cs.iusb.edu/~vasile/logic/indexLF.html>
- The above items were successfully used to run the First International Competition on Logic Form Identification with 27 enrolled participants from all over the world including research teams from academia and industry such as: MITRE, The University of Sidney, Stanford University, University of Berkeley, Language Computer Corporation, etc.
- Students were courted to get involved in the project

Log of activities during the funded period:

November-December 2003

- prepared and post the guidelines of the competition and posted at <http://www.cs.iusb.edu/~vasile/logic/indexLF.html>
- opened a discussion forum on Yahoo groups, called logicforms and coordinated discussion among active participants. The discussion forum was created in December 18 2003 and it is since active. More information is available here: <http://groups.yahoo.com/group/logicforms/>
- selected a representative a set of 500 English sentences and 500 concept definitions from WordNet

January-February 2004

- on Jan 15 the guidelines and trial data were made available to participants
- tagged each individual word with a part of speech tag using Brill's tagger
- parsed each sentence using Collins' parser
- ran PI's LF software (developed for a particular task and thus not perfectly suitable for open text sentences) on parsed sentences
- double checked the output of the previous activity
- continuously monitored and mediated the discussion forum on Yahoo

March-April 2004

- coordinated with all participants the best time frame for distribute the test data (set of English sentences) and collect the response (the sentences in Logic Form) after one week
- The test week was: **March 25 - March 31**
- The software evaluator was developed
- In April the scoring of outputs submitted by participants were evaluated and the report and a paper was submitted for peer-review to SENSEVAL organizers (www.senseval.org) which coordinated efforts for many language evaluation tasks.

June-July 2004

- studied the new proposed metrics at: Precision and Recall
- refined the software evaluator
- refined the paper outlining the experience gained, the results obtained by each individual team and communicate the final ranking and the winner and submit the paper for publication
- the paper and participants presented results at the Association of Computational Linguistics (ACL) 2004 international meeting in July 21-26, 2004

July-December 2004

- applied the software evaluator to a large collection of test data and the new results were added to the ACL paper for submission to journal publication

- integrate the work in textual entailment systems that will be the foundation for external funding efforts

January 2005

- prepared this report

The PI is grateful to Erika Zynda for her professionalism and to the Research Committee and IUSB for funding this work. I also thank Dr. Hakimzadeh, the chair of the computer science department, for actively supporting the project by providing the necessary computing resources to successfully complete the work. Some of the colleagues in the department were also helpful including Dr. William Knight and Dr. Ruth Schwarz.