

Coling 2008

**22nd International Conference on
Computational Linguistics**

**Educational Natural Language Processing
Tutorial notes**

Iryna Gurevych and Delphine Bernhard
Ubiquitous Knowledge Processing Lab
Technische Universität Darmstadt

17 August 2008
Manchester, UK

©2008, Iryna Gurevych and Delphine Bernhard

Licensed to the Coling 2008 Organizing Committee for publication in Coling 2008 and for re-publishing in any form or medium.

All rights reserved.

Order copies of this and other Coling proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

Design by Chimney Design, Brighton, UK

Production and manufacture by One Digital, Brighton, UK

Tutorial instructors

Iryna Gurevych (gurevych@tk.informatik.tu-darmstadt.de)

Iryna Gurevych is head of the Ubiquitous Knowledge Processing (UKP) Lab at the University of Darmstadt. Her research in recent years has focused on the application of lexical semantic knowledge in such areas of NLP as spoken dialogue summarization, information retrieval, and question answering for educational purposes, e.g. electronic career guidance, or question answering based on question-answer repositories in Web 2.0 applied to e-Learning. Her areas of expertise include algorithms for computational lexical semantics and user-generated discourse processing. She guided the development of the high-performance Java-based Wikipedia and Wiktionary APIs as well as projects in collaborative annotation, information filtering and sentiment analysis for eLearning.

Delphine Bernhard(delphine@tk.informatik.tu-darmstadt.de)

Delphine Bernhard is a senior researcher in the Ubiquitous Knowledge Processing Lab at the University of Darmstadt. She obtained her PhD in 2006 from the Université de Grenoble I, where she worked on terminology extraction from domain specific texts and unsupervised morphological analysis. Her current work focuses on the improvement of question answering to meet the specific needs of learners. Her further research projects include NLP for user generated discourse and quality assessment of social media contents.

Outline

Typical Web 2.0 tools such as wikis, blogs, and podcasts have recently entered the classroom and foster interactions between learners and tutors, within the new eLearning 2.0 paradigm. As a result, eLearning 2.0 makes large amounts of eLearning discourse available for NLP within the field of research that we call "Educational Natural Language Processing" (e-NLP). Research on e-NLP has existed for a long time and has focused on e.g. automatic dialogue-based tutoring systems (Litman & Forbes-Riley, 2006), or essay scoring (Attali & Burstein, 2006). Moreover, several workshops on "Building Educational Applications Using NLP" and related topics have already taken place at major NLP conferences, such as HLT-NAACL 2003, COLING 2004, ACL 2005 and ACL 2008.

Educational applications are particularly challenging for NLP since they require an adaptation and practical application of NLP techniques to various types of discourse, e.g. tutoring dialogues which are different from typical task-oriented spoken dialogue systems. Moreover, educational applications place strong requirements on NLP systems, which have to be robust yet accurate. Therefore, this is an important application domain and a source of innovation for NLP as a field, as shown by recent works from Feng et al. (2006), Kim et al. (2006), Malioutov & Barzilay (2006), Mihalcea & Csomai (2007), to name just a few.

In this tutorial, we will review a variety of uses of NLP in the educational domain and point to emerging trends which call for new types of applications. The tutorial will be relevant to a broad audience of NLP researchers interested in applying NLP techniques to new challenging domains, such as eLearning.

Contents

Slide numbers

| | | |
|----|--|----------|
| 1. | Introduction: eLearning and NLP..... | 7-12 |
| 2. | Automatic generation of exercises..... | 13-42 |
| | a) Computer-based testing..... | 13-18 |
| | b) Multiple-choice questions..... | 19-22 |
| | c) Fill-in-the-blank questions..... | 23-26 |
| | d) Multiple-choice cloze questions..... | 27-31 |
| | e) Matching test items..... | 32-34 |
| | f) Error correction questions..... | 35 |
| | g) Evaluation..... | 36-42 |
| 3. | Assessment of learner generated discourse..... | 43-90 |
| | a) Essay scoring..... | 43-64 |
| | b) Plagiarism..... | 65-79 |
| | c) Short answer assessment..... | 80-85 |
| | d) Speech assessment..... | 86-90 |
| 4. | Reading and writing assistance..... | 91-129 |
| | a) Text readability..... | 91-95 |
| | b) Document retrieval for reading practice..... | 96-99 |
| | c) Text simplification..... | 100-101 |
| | d) Vocabulary assistance..... | 102-113 |
| | e) Spell checking..... | 114-120 |
| | f) Grammar checking..... | 121-125 |
| | g) Dictionary lookup..... | 126-129 |
| 5. | Tutoring systems..... | 130-161 |
| | a) Overview..... | 130-136 |
| | b) Examples: AutoTutor/Why2/ITSpoke..... | 137-161 |
| 6. | Web 2.0 and computer supported collaborative learning..... | 162-187 |
| | a) Web 2.0 & eLearning 2.0..... | 162-171 |
| | b) Quality of user-generated content..... | 172-187 |
| 7. | Example e-NLP application: electronic career guidance..... | 188-202 |
| 8. | Conclusions..... | 203-206 |
| 9. | Bibliography..... | Appendix |

