

EE Department Seminar

April 15, 2011, Friday, 3 p.m.

Yorgo I Stefanopulos Meeting Lounge (KB 217)

Language Technology in Action: Computer-aided Generation of Multiple-Choice Tests

Prof. Dr. Ruslan Mitkov

*Professor of Computational Linguistics and Language Engineering
University of Wolverhampton – United Kingdom*

Natural Language Processing applications are far from perfect and computer-aided language processing with the user in control, is often regarded as an alternative and practical Language Technology approach with a view to delivering real-world implementations.

This talk will present a computer-aided environment (tool) for generation of multiple-choice tests which can operate both in fully automatic and semi-automatic mode where the user is offered the option to post-edit the automatically generated test items. Given the time-consuming and labour-intensive task of developing multiple-choice tests manually, this environment has proven to be of considerable help to test developers (e.g. school teachers, university lecturers, company instructors).

The presentation will introduce the underpinning original methodology for generation of multiple-choice tests from electronic documents. The tool based on this methodology reads the texts, identifies the important concepts and deems which sentences featuring such concepts can give rise to meaningful multiple-choice questions. Questions focusing on the important concepts are generated using transformational grammar rules and distractors are selected by identifying concepts which are semantically close to the correct answer. In addition to employing various NLP techniques including (but not limited to) term extraction and shallow parsing, the tool makes use of language resources such as corpora, Wikipedia and WordNet.

The evaluation carried out shows that whereas the quality of the test items generated with the help of the tool (semi-automatic mode) is not compromised, much time is saved by using the tool to develop such items.

The presentation will conclude with a recent study which investigates which strategy/metrics for selection of distractors based on semantic similarity, offers best results.

Prof. Dr. Ruslan Mitkov has been working in Natural Language Processing (NLP), Computational Linguistics, Corpus Linguistics, Machine Translation, Translation Technology and related areas since the early 1980s. His research output was highlighted as being internationally leading in the last UK Research Assessment Exercise (RAE 2008). Whereas Prof. Mitkov is best known for his seminal contributions to the areas of anaphora resolution and automatic generation of multiple-choice tests, his extensively cited research (more than 160 publications including 9 books, 21 journal articles and 20 book chapters) also covers topics such as machine translation, natural language generation, automatic summarisation, computer-aided language processing, centering, translation memory, evaluation, corpus annotation,

bilingual term extraction, automatic identification of cognates and false friends, and NLP-driven corpus-based study of translation universals. Mitkov is author of the monograph *Anaphora resolution* (Longman) and sole Editor of *The Oxford Handbook of Computational Linguistics* (Oxford University Press). Current prestigious projects include his role as Executive Editor of the Journal of Natural Language Engineering (Cambridge University Press), Editor-in-Chief of the Natural Language Processing book series of John Benjamins publishers, and Consulting Editor of Oxford University Press publications in Computational Linguistics. He is also working on the forthcoming Oxford Dictionary of Computational Linguistics (co-authored with Patrick Hanks) and the forthcoming second, substantially revised edition of the Oxford Handbook of Computational Linguistics. Prof. Mitkov has been invited as a keynote speaker at a number of international conferences. He has acted as Programme Chair of various renowned international conferences on Natural Language Processing (NLP), Machine Translation, Translation Technology, Corpus Linguistics and Anaphora Resolution among which RANLP and the annual Aslib Translating and the Computer London conference. He is asked on a regular basis to review for leading international funding bodies and organisations in addition to professorial applications/assessments both in North America and Europe. Ruslan Mitkov is regularly asked to review for leading journals, publishers and conferences and serve as a member of Programme Committees or Editorial Boards. Prof. Mitkov has been an external examiner of many doctoral theses and curricula in the UK and abroad, including Master's programmes related to NLP and Translation Technology. Prof. Mitkov has considerable external funding to his credit and is currently managing several large projects, some of which are funded by UK research councils, by the EC as well as by companies and users from the UK and USA. Ruslan Mitkov received his MSc from the Humboldt University in Berlin, his PhD from the Technical University in Dresden and worked as a Research Professor at the Institute of Mathematics, Bulgarian Academy of Sciences, Sofia. Mitkov is Professor of Computational Linguistics and Language Engineering at the School of Humanities, Languages and Social Sciences at the University of Wolverhampton which he joined in 1995 and where he set up the Research Group in Computational Linguistics. His Research Group has emerged as an internationally leading unit in applied Natural Language Processing (the output of the Research Group was classed by RAE 2008 as internationally leading, internationally excellent and internationally recognised) and members of the group have won awards in different NLP/shared-task competitions. In addition to being Head of the Research Group in Computational Linguistics, Prof. Mitkov is also Director of the Research Institute in Information and Language Processing. The Research Institute consists of the Research Group in Computational Linguistics and the Research Group in Statistical Cybermetrics, which is another top performer in the recent RAE.