

Plenary Lecturer I

Discovering semantic links in texts. Corpora and projects

Dan Cristea

„Alexandru Ioan Cuza” University of Iași
Faculty of Computer Science
Iași, ROMANIA
E-mail: danu.cristea@gmail.com

Free texts are extremely rich in semantic relations, which we, human beings, are at ease to decipher while reading. If this would not be the case, important messages contained there would be obscure and texts would be incomprehensible. In order to make the machine reach a human-like level of performance, properly annotated corpora should be built. In the first part of my talk I will present an experiment developed with our students in Computational Linguistics to acquire a corpus of entities and semantic relations, the QuoVadis project. In the second part of the talk, I will bring forward another human capacity, which we would like to reproduce on machine, the ability to make connections between mentions of notorious entities in books and their virtual realizations. The MappingBooks project, which I will shortly describe, realizes a technology that transforms the classical book onto a multi-dimensional mash-up which combine textual, geographical and web-found data with personal information related to the reader.

Brief Biography of the Speaker: Dan Cristea is a professor at the „Alexandru Ioan Cuza” University of Iași (UAIC), Faculty of Computer Science (FII), the director of the Research Department in FII. He also holds a part-time position in the Institute of Computer Science of the Romanian Academy, the Iasi branch, and is a correspondent member of the Romanian Academy. Back in the ‘80s he has initiated a line of research in Computational Linguistics and Natural Language Processing in UAIC, which has grown during the years both numerically and qualitatively, up to the actual NLP-Group@UAIC-FII. He is the initiator of the series of EUROLAN Summer Schools in Natural Language Processing (started in 1993, with its 12th edition, in July 2015, hosted by the University “Lucian Blaga” of Sibiu) and a co-director of the series of conferences dedicated to resources and tools for processing the Romanian language (the ConsILR conferences, with its 11th edition – in November this year, in Iasi). He is known for his work on discourse related topics (such as veins theory and anaphora resolution), hierarchical description of linguistic metadata, multilingual linguistic workflows, computational lexicography, and has contributions in the creation of electronic resources for Romanian language (among which, the Romanian WordNet, the electronic version of the Thesaurus Dictionary of the Romanian Language, and the Computational Representative Corpus of Contemporary Romanian Language).