

MADAS Master in Data Science for Complex Economic Systems

**MSS - Madas Seminars Series
in Data Sciences and Complexity 2017-2018**

SEMANTIC RESOURCES FOR TEXT DOCUMENTS PROCESSING

Daniele Radicioni, Enrico Mensa and Davide Colla
University of Torino, Department of Computer Science

14 December 2017 @ 14:00

The growth of the Web and the tremendous spread of social networks exert a strong pressure on computational linguistics in areas as diverse as documents categorization, conceptual categorization, automatic text summarization, question answering, and many other linked tasks. Industry demands tools to handle documents, to extract information, unveil trends and to grasp documents semantics; these are widely acknowledged to be challenging, open research problems. One chief question underlying the whole talk is thus whether and to what extent it is possible to automatically characterize text documents semantics. We will briefly survey theories and approaches to conceptual categorization, and introduce DUAL-PECCS, a Knowledge Representation and Reasoning system. DUAL-PECCS solves simple riddles by pairing logical inference (which is performed on formal ontologies) and common-sense reasoning (which is performed based on vectorial representations). Relatedly, we will show in how far common-sense knowledge is necessary to perform even the easiest tasks, and introduce COVER, a recently developed lexical resource that enables systems to perform approximate, non-monotonic reasoning by combining lexicographic precision and common sense knowledge. Finally, we will point out applications in some specific tasks, such as keywords extraction, question answering, conceptual similarity: we will present and discuss our experimental results, and compare them to those in literature.

DANIELE P. RADICIONI is researcher and assistant professor at the Department of Computer Science of the University of Turin. He is active in the following research areas: Natural Language Processing, Knowledge Representation and Engineering (with special focus on formal ontologies and conceptual spaces), Text Mining and Information Extraction, Machine Learning for sequential data. Daniele authored over 70 peer reviewed works that appear in journals and conferences relevant to the research in Artificial Intelligence, and participated in various research projects, both at national and international level. His research is currently focused on building systems and lexical resources that enable NLP to attain some traits of the so-called human level intelligence.

ENRICO MENSA is PhD Student at the Department of Computer Science in Torino. Since his Master's Thesis work, he has been involved in the creation of lexical resources and in the study of NLP at large. His efforts led to creation of COVER, a novel semantic resource that provides conceptual common-sense information suitable for a wide range of tasks, such as conceptual similarity, keywords extraction and question answering (<http://nl.di.unito.it/resources/cover/>). His research is now focused on the study of verbal semantics and its impact on language processing.

DAVIDE COLLA holds a MSc degree in Computer Science (summa cum laude) from the University of Turin. He is working as a research fellow on a project carried out at the Computer Science Department on event extraction from text documents. His main interests are in algorithms for Word Sense Disambiguation, semantic coherence measures for text segmentation and keywords extraction, and in the design of semantic similarity measures for text mining.

More information: massimiliano.nuccio@unito.it - www.madas.carloalberto.org

Collegio Carlo Alberto, piazza Arbarello 8, Torino



UNIVERSITÀ DEGLI STUDI DI TORINO