

GET 2014:

The first Event in Trento

ON MAY 5, 2014

ABSTRACT

The first Google Event in Trento illustrates collaborations on the research area of Natural Language Processing (NLP) between the researchers from the region and from Google. It also provides some hints of the trendy topics of the company presented by two of their representatives. A round table with the event speakers will offer the possibility to discuss further development and directions of NLP technology from both industrial and academic viewpoints.

PROGRAM

- | | |
|---------------|--|
| 09:00 — 09:05 | Welcome and introduction to the event |
| 09:05 — 09:25 | <i>Practical Structure Prediction: reranking NLP system output</i>
Alessandro Moschitti, University of Trento |
| 09:30 — 10:00 | <i>User demographics and language on YouTube</i>
Katja Filippova, Google Inc. |
| 10:05 — 10:25 | <i>Structural Models for Social Media processing</i>
Aliaksei Severyn, University of Trento |
| 10:30 — 11:00 | <i>Coffee Break</i> |
| 11:00 — 11:30 | <i>Heady: Unsupervised, schema-independent, open-domain event modeling</i>
Daniele Pighin, Google Inc. |
| 11:35 — 11:55 | <i>Your first encounter with a wampimuk: Cross-modal mapping of words and things</i>
Marco Baroni, University of Trento |
| 12:00 — 12:20 | <i>Computational explorations of creative language</i>
Carlo Strapparava, FBK |
| 12:25 — 13:00 | <i>Panel with the speakers</i> |

TALK INFORMATION

Title: *Practical Structured Prediction: reranking NLP system output*

Speaker: Alessandro Moschitti, University of Trento

Abstract. Almost all Natural Language Processing (NLP) applications deal with syntactic and semantic structures, requiring therefore the use of models for processing structured input and generating structured output. Given the complexity of natural language, heuristic methods demonstrate often to be inadequate for building NLP systems and also require a considerable design effort. Statistical machine learning methods for structured prediction are an effective alternative to the approach above. However, to our knowledge, the only efficient algorithms enabling the use of structures in both the input (in the form of structural kernels) and the output of a system consist in two steps: (i) the generation of a list of hypotheses of the structured output and (ii) the application of *learning to rank* approaches to select the best hypothesis of the list. This talk will show methods for practical linguistic structure prediction for several applications, ranging from, concept segmentation and labeling to named entity recognition and opinion mining.

Bio: Alessandro Moschitti is a professor of the Department of Computer Science and Information Engineering of the University of Trento. He obtained his PhD in Computer Science from the University of Rome in 2003. He has worked as an associate researcher for the University of Texas at Dallas, as a visiting professor for the Universities of Columbia (NY), Colorado and John Hopkins and as visiting researcher for the IBM Watson, NY, during and after the *Jeopardy! Grand Challenge*. He has coordinated several USA and European projects. Currently, he is leading an international project with MIT and the Qatar Computing Research Institute. He has devised innovative kernels for advanced syntactic and semantic processing for support vector and other kernel-based machines, published in more than 180 scientific articles. He has been an active PC member of several research communities. He is the General Chair of EMNLP 2014 and on the editorial board of several journals, e.g., JAIR, JNLE and JoDS. He has received three IBM Faculty Awards, one **Google Faculty Award** and several best paper awards.

Title: *User demographics and language on YouTube*

Speaker: Katja Filippova, Google Inc.

Abstract. YouTube is one of the most popular web-sites in the world and attracts masses of users who not only watch but also engage in conversations by

commenting on the videos. These users have very diverse interests and backgrounds, reflected in their comments and language. In this talk I will give a few examples of how a statistical analysis of the comments provides us with sociolinguistic insights and may help understand the user demographics beyond what the users list in their public profiles.

Bio: Katja Filippova received her PhD from the Darmstadt University of Technology in 2009. In 2006-2009 she was a research assistant in the NLP group at European Media Lab in Heidelberg (now called the Heidelberg Institute for Theoretical Studies) working on sentence fusion and compression. In 2009-2011 she was a postdoc and since May 2011 a research scientist at Google. Her research interests include statistical approaches to text-to-text generation and making use of user-generated content.

Title: *Structural Models for Social Media processing*

Speaker: Aliaksei Severyn, University of Trento

Abstract: Social media such as Twitter, Facebook or YouTube contain rapidly changing information generated by millions of users that can dramatically affect the reputation of a person or an organization. This raises the importance of automatic extraction of information and sentiment expressed in social media. Previous approaches primarily based on bag-of-words (BOW) features and features extracted from sentiment dictionaries have been shown to produce rather accurate models that are hard to beat by more involved models. In this talk we demonstrate an approach where structural kernels, e.g., tree kernels, are used to handle automatic feature engineering to augment the conventional BOW models. Our empirical results show that the structural features automatically extracted and learned by our kernel-learning framework provide an orthogonal source of information to the conventional BOW models also resulting in more robust and adaptable models. We exemplify our approach on two applications: Opinion Mining on YouTube comments and retrieval of tweets from TREC Microblog challenge.

Bio: Aliaksei Severyn is currently pursuing a Phd in Computer Science advised by Alessandro Moschitti. His primary focus is on applied Machine Learning and Natural Language Processing with applications areas including Semantic Role Labeling, Question Answering, Semantic Textual Similarity and Sentiment Analysis. He has interned at Google where he worked within the NLP team. More recently, he got the **Google Europe Doctoral Fellowship Award** in Machine Learning 2013. He published in top-tier NLP and IR conferences such as ACL, EMNLP, CoNLL, SIGIR, CIKM and at the European Conference of Machine Learning, from which he received the Best Student Paper award in 2011.

Title: *Heady: Unsupervised, schema-independent, open-domain event modeling*

Speaker: Daniele Pighin, Google Inc.

Abstract: Understanding what a piece of text is talking about is a key requirement to be able to access information semantically and to keep world models in sync with reality. In this talk I will briefly introduce HEADY, a completely unsupervised, open-domain, schema-independent framework for modeling and learning events. Heady can be used to generate abstractive headlines for unseen news, or to update existing knowledge bases when a change in the state of the world is detected. After sketching Heady's architecture, I will present the different methods that we use to extract textual event representations, which are also grammatical sentences, and discuss some of the open problems that we are currently trying to solve.

Bio: Daniele got his Ph.D. from the University of Trento (Italy) with a thesis on "Greedy Feature Selection in Tree Kernel Spaces". During the Ph.D. years he has worked mostly on Semantic Role Labeling (for English and Arabic) and Relation Extraction problems, with brief incursions into Machine Translation territory. Before joining Google (first as postdoc, then as research scientist), he spent two years in Barcelona (Spain) working on the slippery topic of Quality Estimation for Machine Translation. He is fascinated by the facets and subtleties of Natural Language. He loves building NLP systems and applications to distill relevant information from noisy text corpora. At Google he is very happy because he has never seen so many words to play with in the same place.

Title: *Your first encounter with a wampimuk: Cross-modal mapping of words and things*

Speaker: Marco Baroni, University of Trento

Abstract: Vectors recording basic statistical properties of objects of interest are popular means of representation both in computational linguistics (where they are used to capture semantic properties of words, among others) and in computer vision (where "bags-of-visual-words" vectors are a popular method to encode images for retrieval and recognition tasks). In my talk, I will present a number of experiments in which vectors are mapped across the textual and visual modalities, in order to establish a link between words and what they denote. I will discuss both the practical and the cognitive implications of these experiments.

Bio: Marco Baroni received a PhD in linguistics from UCLA in 2000. After various experiences in academia and industry, he is currently associate professor at the Center for Mind and Brain Sciences of the University of Trento. Marco's research focuses on statistical methods to extract semantic and conceptual knowledge from large collections of text and images, and on the implications that these computational simulations have for our understanding of human semantic processing. To pursue these topics, he was awarded a **Google Research Award** in 2010, and an ERC Starting Grant in 2011.

Title: *Computational explorations of creative language*

Speaker: Carlo Strapparava, Fondazione Bruno Kessler

Abstract: Dealing with creative language and in particular with affective, persuasive and even humorous language has often been considered outside the scope of computational linguistics. Nonetheless, it is possible to exploit current NLP techniques starting some explorations about it. We briefly review some computational experiences about these typical creative genres.

Bio: Carlo Strapparava is a senior researcher at FBK-irst in the Human Language Technologies Unit. His research activity covers artificial intelligence, natural language processing, cognitive science, knowledge-based systems, user models, adaptive hypermedia, lexical knowledge bases, word-sense disambiguation, affective computing and computational humor. He is the author of over 170 papers, published in scientific journals, book chapters and in conference proceedings. On June 2011, he was awarded with a **Google Research Award** on Natural Language Processing, specifically on the computational treatment of affective and creative language.

EVENT DETAILS

When: May 5, 2014

09:00 AM to 01:00 PM

Where: Polo Scientifico e Tecnologico, Aula B105 (Povo 2)

Via Sommarive 9, 38123 Trento - Italy

ORGANIZERS: Alessandro Moschitti and Katja Filippova

For more information, please contact moschitti_a_t@disi.unint.it