

WebNLG 2016

**Proceedings of the
2nd International Workshop
on
Natural Language Generation
and the Semantic Web**

6 September 2016
Edinburgh, Scotland

WebNLG 2016 is sponsored by:

the French National Research Agency Project ANR-14-CE24-0033 “Generating Text from Semantic Web Data” (WebNLG)



©2016 The Association for Computational Linguistics

Order copies of this and other ACL 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

Introduction

It is with great pleasure that we present the current volume of papers accepted for presentation at the 2nd International Workshop on Natural Language Generation and the Semantic Web to be held in Edinburgh, Scotland on September 6th, 2016.

The WebNLG 2016 workshop is a follow up to a first WebNLG workshop which was held in Nancy on June 12th, 2015. Funded by the French ANR WebNLG Project, these two workshops aim to provide a forum for presenting and discussing research on Natural Language Generation from Semantic Web data.

WebNLG 2016 invited submissions on all topics related to Natural Language Generation and the Semantic Web. We received 14 submissions from all over the world. Of these 5 long papers and 8 short papers were accepted for presentation. The long papers will be presented orally, and the short papers as posters.

In addition, WebNLG 2016 hosts an Invited talk by Roberto Navigli from Sapienza University (Rome, Italy) on the past, present and future of Babelnet .

We are indebted to the authors and to the members of our program committee for their work which contributed to make for a very enjoyable workshop. We are also delighted that Roberto Navigli agreed to give an invited talk at WebNLG 2016. Last but not least, many thanks go to the local organisation team, Emilie Colin, Bikash Gyawali, Mariem Mahfoudh and Laura Perez-Beltrachini for handling the website and the preparation of the meeting.

Claire Gardent and Aldo Gangemi
Program co-Chairs for WebNLG 2016

Program Chairs:

Aldo Gangemi, Université Paris 13, Paris (France)

Claire Gardent, CNRS/LORIA, Nancy (France)

Organisation Committee:

Bikash Gyawali CNRS/LORIA, Nancy (France). Chair.

Laura Perez-Beltrachini CNRS/LORIA, Nancy (France), Chair.

Émilie Colin, CNRS/LORIA, Nancy (France), Webmaster.

Mariem Mahfoudh, CNRS/LORIA, Nancy (France)

Program Committee:

Mehwish Alam, LIPN Université Paris 13 (France)

Nathalie Aussenac-Gilles, CNRS/IRIT Toulouse (France)

Valerio Basile, INRIA, Sophia Antipolis (France)

Gerard Casamayor, Universitat Pompeu Fabra (Spain)

Vinay Chaudhri, SRI International, Menlo Park (USA)

Mathieu Dacquin, The Open University (UK)

Claudia d'Amato, Bari University (Italy)

Brian Davis, INSIGHT, Galway (Ireland)

Marc Dymetman, XRCE, Grenoble (France)

Enrico Franconi, KRDB, Bolzano (Italy)

Bikash Gyawali, CNRS/LORIA, Nancy (France)

Guy Lapalme, RALI / Université de Montréal (Canada)

Shao-Fen Liang, Kings College London (UK)

Elena Lloret, University of Alicante (Spain)

Vanessa Lopez, IBM Ireland Research Lab, Dublin (Ireland)

Mariem Mahfoudh, CNRS/LORIA, Nancy (France)

Yassine M'rabet, U.S. National Library of Medicine, Bethesda (USA)

Shashi Narayan, University of Edinburgh (UK)

Axel-Cyrille Ngonga-Ngomo, University of Leipzig (Germany)

Laura Perez-Beltrachini, CNRS/LORIA, Nancy (France)

Sergio Tessaris, KRDB, Bolzano (Italy)

Allan Third, The Open University (UK)

Yannick Toussaint, INRIA/LORIA, Nancy (France)

Christina Unger, CITEC, Universitt Bielefeld (Germany)

Invited Speaker

Roberto Navigli, Sapienza Università di Roma, Roma, Italy

BabelNet: past, present and future

In this talk I will overview work done in my group at the Linguistic Computing Laboratory in the Computer Science Department of the Sapienza University of Rome which addresses key problems in multilingual lexical semantics. I will start from a brief introduction to BabelNet, the largest multilingual semantic network and encyclopedic dictionary covering 14 million concepts and entities, and 271 languages, also at the core of the so-called Linguistic Linked Open Data cloud. I will move on to Word Sense Disambiguation and Entity Linking in arbitrary languages with "zero training" (Babelify) and then present recent latent and explicit vector representations of meaning which obtain state-of-the-art results in several NLP tasks. Finally, I will present my plan for making BabelNet a sustainable, continuously-improved resource. This is joint work with several people from my NLP group at Sapienza.

Roberto Navigli is an Associate Professor in the Department of Computer Science of the Sapienza University of Rome. He was awarded the Marco Somalvico 2013 AI*IA Prize for the best young researcher in AI. He was the first Italian recipient of an ERC Starting Grant in computer science (2011-2016), and a co-PI of a Google Focused Research Award on Natural Language Understanding. In 2015 he received the META prize for groundbreaking work in overcoming language barriers with BabelNet. His research lies in the field of multilingual Natural Language Processing. Currently he is an Associate Editor of the Artificial Intelligence Journal.

Table of Contents

<i>Generating Sets of Related Sentences from Input Seed Features</i> Cristina Barros and Elena Lloret	1
<i>A Repository of Frame Instance Lexicalizations for Generation</i> Basile Valerio	5
<i>Processing Document Collections to Automatically Extract Linked Data: Semantic Storytelling Technologies for Smart Curation Workflows</i> Peter Bourgonje, Julian Moreno Schneider, Georg Rehm and Felix Sasaki	13
<i>On the Robustness of Standalone Referring Expression Generation Algorithms Using RDF Data</i> Pablo Ariel Duboue, Martin Ariel Domínguez and Paula Estrella	17
<i>Content Selection through Paraphrase Detection: Capturing different Semantic Realisations of the Same Idea</i> Elena Lloret and Claire Gardent	25
<i>Aligning Texts and Knowledge Bases with Semantic Sentence Simplification</i> Yassine Mrabet, Pavlos Vougiouklis, Halil Kilicoglu, Claire Gardent, Dina Demner-Fushman, Jonathon Hare and Elena Simperl	29
<i>Building a System for Stock News Generation in Russian</i> Liubov Nesterenko	37
<i>Content Selection as Semantic-Based Ontology Exploration</i> Laura Perez-Beltrachini, Claire Gardent, Anselme Revuz and Saptarashmi Bandyopadhyay	41
<i>README Generation from an OWL Ontology Describing NLP Tools</i> Driss Sadoun, Satenik Mkhitarian, Damien Nouvel and Mathieu Valette	46
<i>Comparing the Template-Based Approach to GF: the case of Afrikaans</i> Lauren Sanby, Ion Todd and C.Maria Keet	50
<i>Generating Paraphrases from DBpedia using Deep Learning</i> Amin Sleimi and Claire Gardent	54
<i>Automatic Tweet Generation From Traffic Incident Data</i> Khoa Tran and Fred Popowich	59
<i>Analysing the Integration of Semantic Web Features for Document Planning across Genres</i> Marta Vicente and Elena Lloret	67

Conference Program

Tuesday, September 6, 2016

9:00–10:00 Invited Talk

BabelNet: Past, Present and Future
Roberto Navigli

10:00–10:45 **Posters**

Analysing the Integration of Semantic Web Features for Document Planning across Genres

Marta Vicente and Elena Lloret

Building a System for Stock News Generation in Russian

Liubov Nesterenko

Comparing the Template-Based Approach to GF: the case of Afrikaans

Lauren Sanby, Ion Todd and C. Maria Keet

Content Selection as Semantic-Based Ontology Exploration

Laura Perez-Beltrachini, Claire Gardent, Anselme Revuz and Saptarashmi Bandyopadhyay

Content Selection through Paraphrase Detection: Capturing different Semantic Realisations of the Same Idea

Elena Lloret and Claire Gardent

Generating Sets of Related Sentences from Input Seed Features

Cristina Barros and Elena Lloret

Processing Document Collections to Automatically Extract Linked Data: Semantic Storytelling Technologies for Smart Curation Workflows

Peter Bourgonje, Julian Moreno Schneider, Georg Rehm and Felix Sasaki

ReadME Generation from an OWL Ontology Describing NLP Tools

Driss Sadoun, Satenik Mkhitarian, Damien Nouvel and Mathieu Valette

Oral presentations

10:45–11:05 *Generating Paraphrases from DBPedia using Deep Learning*

Amin Sleimi and Claire Gardent

11:05–11:25 *Aligning Texts and Knowledge Bases with Semantic Sentence Simplification*

Yassine Mrabet, Pavlos Vougiouklis, Halil Kilicoglu, Claire Gardent, Dina Demner-Fushman, Jonathon Hare and Elena Simperl

Oral presentations (continued)

- 11:25–11:45 *On the Robustness of Standalone Referring Expression Generation Algorithms Using RDF Data*
Pablo Ariel Duboue, Martin Ariel Domínguez and Paula Estrella
- 11:45–12:05 *A Repository of Frame Instance Lexicalizations for Generation*
Basile Valerio
- 12:05–12:25 *Automatic Tweet Generation From Traffic Incident Data*
Khoa Tran and Fred Popowitch