

# Computational Linguistics in Bulgaria



## CLIB 2014

4 September, 2014  
Sofia, Bulgaria

CLIB 2014 is organised within the project  
*Integrating New Practices and Knowledge in  
Undergraduate and Graduate Courses in  
Computational Linguistics*  
Grant № BG051PO001-3.3.06-0022.



European Union



European Social Fund

The project is implemented with the financial support of the *Human Resources Development Operational Programme 2007-2013* co-financed by the European Social Fund of the European Union.

**CLIB 2014 is organised by:**

**The Department of Computational Linguistics,  
Institute for Bulgarian Language,  
Bulgarian Academy of Sciences**

**The Faculty of Slavic Studies,**

**The Faculty of Mathematics and Informatics,  
Sofia University**

**Publication and cataloguing information**

Title: **Proceedings of the First International Conference  
*Computational Linguistics in Bulgaria***

ISSN: **2367-5578**

Edited by: **Svetlozara Leseva, Tsvetana Dimitrova, Ivelina Stoyanova,  
Ekaterina Tarpomanova, Rositsa Dekova**

Published and  
distributed by: **The Institute for Bulgarian Language Prof. Lyubomir Andreychin –  
Bulgarian Academy of Sciences**

Editorial address: **Institute for Bulgarian Language Prof. Lyubomir Andreychin,  
Bulgarian Academy of Sciences**

**52 Shipchenski prohod blvd, bldg. 17  
Sofia 1113, Bulgaria  
+359 2/ 872 23 02**

Copyright of each paper stays with the respective authors.  
The works in the Proceedings are licensed under a Creative Commons Attribution 4.0 International  
Licence (CC BY 4.0). Licence details:  
<http://creativecommons.org/licenses/by/4.0>

Proceedings of the  
First International Conference

*Computational Linguistics in  
Bulgaria*

**CLIB 2014**

4 September 2014  
Sofia, Bulgaria

## PREFACE

We are excited to welcome you to the inaugural edition of the international conference *Computational Linguistics in Bulgaria* (CLIB 2014) in Sofia, Bulgaria!

CLIB is a joint effort launched by the Department of Computational Linguistics (DCL) at the Institute for Bulgarian Language of the Bulgarian Academy of Sciences together with the Faculty of Slavic Studies and the Faculty of Mathematics and Informatics at Sofia University.

CLIB aspires to foster the NLP community in Bulgaria and further the cooperation among researchers working in NLP for Bulgarian around the world. The need for a conference dedicated to NLP research dealing with or applicable to Bulgarian has been felt for quite some time. We believe that building a strong community of researchers and teams who have chosen to work on Bulgarian is a key factor to meeting the challenges and requirements posed to computational linguistics and NLP in Bulgaria. We share the hope that CLIB will establish itself as an international forum for sharing high-quality scientific work in all areas of computational linguistics and NLP and will grow in scope and scale with each new edition. The CLIB community will be dedicated to supporting the creation and improvement of advanced NLP resources, tools and technologies for mono- and multilingual language processing, machine translation and translation aids, content creation, localisation and personalisation, speech recognition and generation, information retrieval and information extraction.

The Conference was made possible due to the hard work of many people. We would like to thank the authors who trusted us and submitted their contributions to CLIB 2014. Their efforts and high-quality research are the chief factor that enabled us to create an interesting and solid scientific programme. We would also like to thank our industrial participants for sharing their insights, ideas and know-how with the research community. Let us also express our sincere gratitude to the members of the Programme Committee, who accepted to join us and invested a lot of expertise to provide valuable feedback to the authors. Special thanks are due to Prof. Svetla Kœva, who is the person behind the whole CLIB concept.

We hope that CLIB 2014 will be a useful and productive experience that we all will enjoy!

**CLIB Organising Committee**

## PROGRAMMING COMMITTEE

**Galia Angelova** – Institute of Information and Communication Technologies,  
Bulgarian Academy of Sciences

**Dan Cristea** – Alexandru Ioan Cuza University of Iași

**Radovan Garabík** – Ludovít Štúr Institute of Linguistics

**Mariana Damova** – Mozaika Ltd., Bulgaria

**Ivan Derzhanski** – Institute of Mathematics and Computer Science, Bulgarian Academy of Sciences

**Kjetil Rå Hauge** – University of Oslo

**Verginica Barbu Mititelu** – Research Institute for Artificial Intelligence, Romanian Academy

**Stoyan Mihov** – Institute of Information and Communication Technologies,  
Bulgarian Academy of Sciences

**Preslav Nakov** – Qatar Computing Research Institute

**Svetla Koeva** – Institute for Bulgarian Language, Bulgarian Academy of Sciences

**Cvetana Krstev** – University of Belgrade

**Éric Laporte** – University of Paris-Est Marne-la-Vallée

**Denis Maurel** – François-Rabelais University of Tours

**Maciej Ogrodniczuk** – Institute of Computer Science, Polish Academy of Sciences

**Karel Oliva** – Institute of the Czech Language, Academy of Sciences of the Czech Republic

**Maciej Piasecki** – Wrocław University of Technology

**Agata Savary** – François-Rabelais University of Tours

**Jan Šnajder** – University of Zagreb

**Ivelina Stoyanova** – Institute for Bulgarian Language, Bulgarian Academy of Sciences

**Marko Tadić** – University of Zagreb

**Hristo Tanev** – Joint Research Centre of the European Commission in Ispra, Italy

**Tinko Tinchev** – Sofia University St. Kliment Ohridski

**Dan Tufiş** – Research Institute for Artificial Intelligence, Romanian Academy

**Duško Vitas** – University of Belgrade

**Radka Vlahova** – Sofia University St. Kliment Ohridski

## ORGANISING COMMITTEE

**Rositsa Dekova** – Institute for Bulgarian Language, Bulgarian Academy of Sciences

**Tsvetana Dimitrova** – Institute for Bulgarian Language, Bulgarian Academy of Sciences

**Svetlozara Leseva** – Institute for Bulgarian Language, Bulgarian Academy of Sciences

**Vladislav Nenchev** – Sofia University St. Kliment Ohridski

**Bilyana Radeva** – Sofia University St. Kliment Ohridski

**Andrey Sariiev** – Sofia University St. Kliment Ohridski

**Ekaterina Tarpomanova** – Sofia University St. Kliment Ohridski

**Stefan Vatev** – Sofia University St. Kliment Ohridski

## INVITED SPEAKERS

**Prof. Cvetana Krstev.** *Developing Resources for the Culinary Domain.*

**Prof. Eric Laporte.** *Interaction between Linguists and Machine Learning.*

# Table of Contents

Ivan Derzhanski, Rositsa Dekova <i>Electronic Language Resources in Teaching Mathematical Linguistics</i> . . . . .	1
Diman Karagiozov <i>Harnessing Language Technologies in Multilingual Information Channeling Services</i> . . . . .	6
Svetlozara Leseva, Ivelina Stoyanova, Borislav Rizov, Maria Todorova, Ekaterina Tarpomanova <i>Automatic Semantic Filtering of Morphosemantic Relations in WordNet</i> . . . . .	14
Ekaterina Tarpomanova, Svetlozara Leseva, Maria Todorova, Tsvetana Dimitrova, Borislav Rizov, Verginica Barbu Mititelu, Elena Irimia <i>Noun-Verb Derivation in the Bulgarian and the Romanian WordNet – A Comparative Approach</i> . .	23
Daniela Majchráková, Ondřej Dušek, Jan Hajič, Agáta Karčová, Radovan Garabík <i>Semi-automatic Detection of Multiword Expressions in the Slovak Dependency Treebank</i> . . . . .	32
Ivelina Stoyanova <i>Automatic Categorisation of Multiword Expressions and Named Entities in Bulgarian</i> . . . . .	40
Ivan Derzhanski, Olena Siruk <i>Temporal Adverbs and Adverbial Expressions in a Corpus of Bulgarian and Ukrainian Parallel Texts</i> . . . . .	49
Tsvetana Dimitrova, Andrej Bojadiev <i>Historical Corpora of Bulgarian Language and Second Position Markers</i> . . . . .	55
Luchezar Jackov <i>Machine Translation Based on WordNet and Dependency Relations</i>	64
Sebastiano Pais, Gael Dias, Rumen Moraliyski <i>Recognize the Generality Relation between sentences using Asymmetric Association Measures</i> . . . . .	73
Sebastiano Pais, Gael Dias, Rumen Moraliyski, Joao Cordeiro <i>Unsupervised and Language-Independent Method to Recognize Textual Entailment by Generality</i>	82