3rd Symposium on Languages, Applications and Technologies

SLATE'14, June 19–20, 2014, Bragança, Portugal

Edited by

Maria João Varanda Pereira
José Paulo Leal
Alberto Simões

Editors

Maria João Varanda Pereira
CCTC
Escola Superior de Tecnologia e Gestão
Instituto Politécnico de Bragança
mjoao@ipb.pt

José Paulo Leal
CRACS & INESC TEC
Faculdade de Ciências
Universidade do Porto
zp@fcc.fc.up.pt

Alberto Simões
CEHUM & CCTC
Instituto de Letras e Ciências Humanas
Universidade do Minho
ambs@ilch.uminho.pt

Funding
This publication is funded by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project PEst-OE/EEI/UI0752/2014.

ACM Classification 1998
D.3 Programming Languages, D.2.12 Interoperability, I.2.7 Natural Language Processing

ISBN 978-3-939897-68-2

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at http://www.dagstuhl.de/dagpub/978-3-939897-68-2.

Publication date
June, 2014

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

License
This work is licensed under a Creative Commons Attribution 3.0 Unported license (CC-BY 3.0): http://creativecommons.org/licenses/by/3.0/legalcode.
In brief, this license authorizes each and everybody to share (to copy, distribute and transmit) the work under the following conditions, without impairing or restricting the authors' moral rights:
- Attribution: The work must be attributed to its authors.

The copyright is retained by the corresponding authors.

Digital Object Identifier: OASIcs.SLATE.2014.i

ISBN 978-3-939897-68-2 **ISSN 2190-6807** **http://www.dagstuhl.de/oasics**

OASIcs – OpenAccess Series in Informatics

OASIcs aims at a suitable publication venue to publish peer-reviewed collections of papers emerging from a scientific event. OASIcs volumes are published according to the principle of Open Access, i.e., they are available online and free of charge.

Editorial Board

- Daniel Cremers (TU München, Germany)
- Barbara Hammer (Universität Bielefeld, Germany)
- Marc Langheinrich (Università della Svizzera Italiana – Lugano, Switzerland)
- Dorothea Wagner (*Editor-in-Chief*, Karlsruher Institut für Technologie, Germany)

ISSN 2190-6807

www.dagstuhl.de/oasics

Contents

Preface
Maria João Varanda Pereira, José Paulo Leal, and *Alberto Simões* viii

Invited Talks

Language-Driven Software Development
José-Luis Sierra 3

An Overview of Open Information Extraction
Pablo Gamallo 13

Program Comprehension

Conclave: Writing Programs to Understand Programs
Nuno Ramos Carvalho, José João Almeida, Maria João Varanda Pereira, and Pedro Rangel Henriques 19

Leveraging Program Comprehension with Concern-oriented Source Code Projections
Jaroslav Porubän and Milan Nosáľ 35

Comment-based Concept Location over System Dependency Graphs
Nuno Pereira, Maria João Varanda Pereira, and Pedro Rangel Henriques 51

Domain Specific Languages

ReCooPLa: a DSL for Coordination-based Reconfiguration of Software Architectures
Flávio Rodrigues, Nuno Oliveira, and Luís S. Barbosa 61

A Workflow Description Language to Orchestrate Multi-Lingual Resources
Rui Brito and José João Almeida 77

Converting Ontologies into DSLs
João M. Sousa Fonseca, Maria João Varanda Pereira, and Pedro Rangel Henriques 85

JSON on Mobile: is there an Efficient Parser?
Ricardo Queirós 93

Unfuzzying Fuzzy Parsing
Pedro Carvalho, Nuno Oliveira, and Pedro Rangel Henriques 101

Programming Languages and Compilers

Contract-Java: Design by Contract in Java with Safe Error Handling
Miguel Oliveira e Silva and Pedro G. Francisco 111

Implementing Python for DrRacket
Pedro Palma Ramos and António Menezes Leitão 127

3^{rd} Symposium on Languages, Applications and Technologies (SLATE'14).
Editors: Maria João Varanda Pereira, José Paulo Leal, and Alberto Simões
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Plagiarism Detection: A Tool Survey and Comparison
Vítor T. Martins, Daniela Fonte, Pedro Rangel Henriques, and Daniela da Cruz .. 143

Target Code Selection by Tilling AST with the Use of Tree Pattern Pushdown Automaton
Jan Janoušek and Jaroslav Málek 159

Semantics in Natural Language Processing

Assigning Polarity Automatically to the Synsets of a Wordnet-like Resource
Hugo Gonçalo Oliveira, António Paulo Santos, and Paulo Gomes 169

Detecting a Tweet's Topic within a Large Number of Portuguese Twitter Trends
Hugo Rosa, João Paulo Carvalho, and Fernando Batista 185

Multiscale Parameter Tuning of a Semantic Relatedness Algorithm
José Paulo Leal and Teresa Costa 201

Rocchio's Model Based on Vector Space Basis Change for Pseudo Relevance Feedback
Rabeb Mbarek, Mohamed Tmar, and Hawete Hattab 215

Automatic Identification of Whole-Part Relations in Portuguese
Ilia Markov, Nuno Mamede, and Jorge Baptista 225

Natural Language Processing Tools and Resources

Automatic Detection of Proverbs and their Variants
Amanda P. Rassi, Jorge Baptista, and Oto Vale 235

Language Identification: a Neural Network Approach
Alberto Simões, José João Almeida, and Simon D. Byers 251

LemPORT: a High-Accuracy Cross-Platform Lemmatizer for Portuguese
Ricardo Rodrigues, Hugo Gonçalo Oliveira, and Paulo Gomes 267

Expanding a Database of Portuguese Tweets
Gaspar Brogueira, Fernando Batista, João P. Carvalho, and Helena Moniz 275

MLT-prealigner: a Tool for Multilingual Text Alignment
Pedro Carvalho and José João Almeida 283

Preface

The communication from man to man evolved, from long time ago to the communication between man and machine. Communication is achieved when the receiver understands the words, the sentences and knows its meaning in a certain context. A successful communication depends on so many factors: the adequacy of the language type (considering the stakeholders), mutual agreement on the language to use, the ability of the issuer to express himself with the proper words and well-constructed sentences, the ability of the receiver to process the information received and react. The communication between man and computer implies preparing the machine with proper software to be able to receive source texts and perform actions. The study of formalisms and the creation of new approaches associated language processing tasks, is an important research topic in the area of Computer Science.

Techniques and approaches have been developed to speed up and make more efficient the use of the languages either improving the processing tasks of well-known programming languages, constructing new program comprehension tools to be used in the maintenance phase, creating domain specific languages or dealing with problems concerning with natural language processing (NLP) and other topics that relate languages with technology. In SLATE 2014 a challenge is proposed to all participants: update the state-of-the-art, discuss solutions for identified problems, present new ideas and have fun.

The symposium is divided in three tracks:

- The HHL (Human-Human Languages) track is concerned with natural language processing issues and their application in several contexts.
- The HCL (Human-Computer Languages) track is dedicated to exchange ideas about language design, processing, assessment and comprehension and an huge number of applications that can be created to deal with this.
- The CCL (Computer-Computer Languages) track whose main goal is to discuss the use and associated technologies of the XML markup language.

This volume contains the proceedings of the 3^{rd} edition of SLATE, held in the School of Technology and Management of Polytechnic Institute of Bragança, Portugal, during 19th–20th June, 2014. This year, SLATE received a total of 20 full paper submissions and 9 short paper submissions. Each submission was reviewed by at least three Program Committee members, from a global group of 63 researchers. At the end of the review process, 12 papers were accepted as full papers, 4 full papers were invited to submit as short papers, and 6 short papers were also accepted for publication and presentation at the symposium. So, SLATE 2014 had a 24% rejection and 22 papers presentations: 12 full papers (20 min + 5 min of questions) and 10 short papers (10 min + 5 min of questions).

This set of presentations is divided into the following five sessions:

Domain Specific Languages, includes one full paper and four short papers dedicated to the creation of new DSLs and techniques to implement this kind of languages.

Programming Languages and Compilers, includes three full papers and one short paper about programming language implementation, source code analysis, and target code generation.

Program Comprehension, includes two full papers and one short paper about different techniques used for program comprehension: identifier analysis, concern-based projections, and system dependency graph enriched with source code comments.

3^{rd} Symposium on Languages, Applications and Technologies (SLATE'14).
Editors: Maria João Varanda Pereira, José Paulo Leal, and Alberto Simões
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Semantics in NLP, includes four full papers and one short paper related to the analysis of semantic in natural language processing, namely on the extraction of semantic relationships from texts, and on the use of semantic-rich structures;

NLP Tools and Resources, includes two full papers and three short papers on identification and analysis of natural language sentences, text alignment and databases.

Moreover, SLATE 2014 program also includes two keynotes: one on **Language-driven Software Development**, by José Luís Sierra from Complutense University of Madrid and another on **Open Information Extraction** by Pablo Gamallo from University of Santiago de Compostela.

The organizers of SLATE 2014 want to thank to many people without whom this event would never be possible: our sponsors Efacec, Computer Science and Technology Center (CCTC), Polytechnic Institute of Bragança (IPB) and Fundação para a Ciência e a Tecnologia (FCT, Portuguese Foundation for Science and Technology); Easychair conference management system; the Program Committee members for spending their time reviewing the papers and writing the reports; the authors of the submitted papers for their contribution and interest in the symposium and, finally, to all participants that came to Bragança to such a fruitful meeting.

Maria João Varanda Pereira
José Paulo Leal
Alberto Simões

List of Authors

José João Almeida
CCTC, Departamento de Informática
Universidade do Minho
Braga, Portugal
jj@di.uminho.pt

Jorge Baptista
INESC-ID Lisboa, L2F
Universidade do Algarve – FCHS/CECL
Faro, Portugal
jbaptis@l2f.inesc-id.pt

Luís S. Barbosa
HASLab – INESC TEC
Universidade do Minho
Braga, Portugal
lsb@di.uminho.pt

Fernando Batista
INESC-ID Lisboa & ISCTE
Instituto Universitário de Lisboa
Lisboa, Portugal
fernando.batista@iscte.pt

Rui Brito
CCTC, Departamento de Informática
Universidade do Minho
Braga, Portugal
ruibrito@di.uminho.pt

Gaspar Brogueira
Laboratório de Sistemas de Língua Falada
INESC-ID, Lisboa, Portugal
gmrba@iscte.pt

Simon D. Byers
AT & T Labs
Bedminster NJ
United States of America
headers@gmail.com

João Paulo Carvalho
INESC-ID Lisboa
IST – Universidade de Lisboa
Lisboa, Portugal
joao.carvalho@inesc-id.pt

Nuno Ramos Carvalho
CCTC, Departamento de Informática
Universidade do Minho
Braga, Portugal
narcarvalho@di.uminho.pt

Pedro Carvalho
Departamento de Informática
Universidade do Minho
Braga, Portugal
pedrocarvalho@di.uminho.pt

Teresa Costa
CRACS & INESC-Porto LA
Faculty of Sciences
University of Porto
Porto, Portugal
teresa.costa@dcc.fc.up.pt

Daniela da Cruz
CCTC, Departamento de Informática
Universidade do Minho
Braga, Portugal
danieladacruz@gmail.com

João Manuel Sousa Fonseca
CCTC, Departamento de Informática
Universidade do Minho
Braga, Portugal
jprophet89@gmail.com

Daniela Fonte
CCTC, Departamento de Informática
Universidade do Minho
Braga, Portugal
danielamoraisfonte@gmail.com

Pedro G. Francisco
University of Aveiro, IEETA
Campus Universitário de Santiago
Aveiro, Portugal
goucha@ua.pt

Pablo Gamallo
Universidade de Santiago de Compostela
Galiza, Spain
pablo.gamallo@usc.es

3^{rd} Symposium on Languages, Applications and Technologies (SLATE'14).
Editors: Maria João Varanda Pereira, José Paulo Leal, and Alberto Simões
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Paulo Gomes
CISUC, Department of Informatics Engineering
University of Coimbra
Coimbra, Portugal
pgomes@dei.uc.pt

Hawete Hattab
Umm Al-qura University
Department of Mathematics
Makkah, KSA
hshattab@uqu.edu.sa

Pedro Rangel Henriques
CCTC, Departamento de Informática
Universidade do Minho
Braga, Portugal
prh@di.uminho.pt

Jan Janoušek
Department of Theoretical Computer Science
Faculty of Information Technologies
Czech Technical University in Prague
Prague, Czech Republic
Jan.Janousek@fit.cvut.cz

José Paulo Leal
CRACS & INESC-Porto LA
Faculty of Sciences
University of Porto
Porto, Portugal
zp@dcc.fc.up.pt

António Menezes Leitão
INESC-ID, Instituto Superior Técnico
Universidade de Lisboa
Lisboa, Portugal
antonio.menezes.leitao@tecnico.ulisboa.pt

Jaroslav Málek
Department of Theoretical Computer Science
Faculty of Information Technologies
Czech Technical University in Prague
Prague, Czech Republic

Nuno Mamede
INESC-ID Lisboa, L2F
Instituto Superior Técnico,
Universidade de Lisboa,
Lisboa, Portugal
Nuno.Mamede@l2f.inesc-id.pt

Ilia Markov
INESC-ID Lisboa, L2F
Universidade do Algarve – FCHS
Faro, Portugal
Ilia.Markov@l2f.inesc-id.pt

Helena Moniz
Laboratório de Sistemas de Língua Falada
INESC-ID, Lisboa, Portugal
helena.moniz@inesc-id.pt

Vítor T. Martins
CCTC, Departamento de Informática
Universidade do Minho
Braga, Portugal
vtiagovm@gmail.com

Rabeb Mbarek
Sfax University
Multimedia Information Systems and Advanced Computing Laboratory
Sfax, Tunisia
rabeb.hattab@gmail.com

Milan Nosáľ
Faculty of Elect. Eng. and Informatics
Technical University of Košice
Košice, Slovakia
milan.nosal@gmail.com

Hugo Gonçalo Oliveira
CISUC, Department of Informatics Engineering
University of Coimbra
Coimbra, Portugal
hroliv@dei.uc.pt

Nuno Oliveira
HASLab – INESC TEC
Universidade do Minho
Braga, Portugal
nunooliveira@di.uminho.pt

Maria João Varanda Pereira
CCTC, Instituto Politécnico de Bragança
Bragança, Portugal
mjoao@ipb.pt

Nuno Pereira
CCTC, Departamento de Informática
Universidade do Minho
Braga, Portugal
nuno.filipe.gomes.pereira@gmail.com

Jaroslav Porubän
Faculty of Elect. Eng. and Informatics
Technical University of Košice
Košice, Slovakia
jaroslav.poruban@tuke.sk

Ricardo Queirós
CRACS & INESC-Porto LA
Escola Superior de Estudos Industriais e de Gestão
Instituto Politécnico do Porto
ricardo.queiros@eu.ipp.pt

Amanda P. Rassi
Federal University of São Carlos-UFSCar
São Carlos, São Paulo, Brasil
aprassi@ualg.pt

Pedro Palma Ramos
INESC-ID, Instituto Superior Técnico
Universidade de Lisboa
Lisboa, Portugal
pedropramos@tecnico.ulisboa.pt

Flávio Rodrigues
HASLab – INESC TEC
Universidade do Minho
Braga, Portugal
pg22826@alunos.uminho.pt

Ricardo Rodrigues
Centre for Informatics and Systems of the University of Coimbra Coimbra, Portugal
rmanuel@dei.uc.pt

Hugo Rosa
INESC-ID Lisboa
IST – Universidade de Lisboa
Lisboa, Portugal
hugohrosa@gmail.com

António Paulo Santos
GECAD, Institute of Engineering
Polytechnic of Porto
Porto, Portugal
pgsa@isep.ipp.pt

Jose-Luis Sierra
Fac. Informática
Universidad Complutense de Madrid
Madrid, Spain
jlsierra@fdi.ucm.es

Miguel Oliveira e Silva
University of Aveiro, IEETA, DETI
Campus Universitário de Santiago
Aveiro, Portugal
mos@ua.pt

Alberto Simões
Centro de Estudos Humanísticos
Universidade do Minho
Braga, Portugal
ambs@ilch.uminho.pt

Mohamed Tmar
Sfax University
Multimedia Information Systems and Advanced Computing Laboratory
Sfax, Tunisia
mohamedtmar@yahoo.fr

Oto Vale
Federal University of São Carlos-UFSCar
São Carlos, São Paulo, Brasil
otovale@ufscar.br

Committees

Program Chairs

Maria João Varanda Pereira
Instituto Politécnico de Bragança, Portugal

José Paulo Leal
Universidade do Porto, Portugal

Alberto Simões
Universidade do Minho, Portugal

Publication Chair

Alberto Simões
Universidade do Minho, Portugal

Program Committee

Salvador Abreu
Universidade de Évora, Portugal

José João Almeida
Universidade do Minho, Portugal

Jorge Baptista
Universidade do Algarve, Portugal

Fernando Batista
ISCTE-IUL & INESC-ID, Portugal

Mario Berón
Universidad Nacional de San Luis, Argentina

Michele Bugliesi
Università Ca'Foscari Venezia, Italy

João M. P. Cardoso
Universidade do Porto & INESC TEC, Portugal

Nuno Ramos Carvalho
Universidade do Minho, Portugal

Matej Crepinsek
Univerza v Mariboru, Slovenia

Daniela da Cruz
Universidade do Minho, Portugal

Jürgen Ebert
Universität Koblenz-Landau, Germany

Gabriel David
Universidade do Porto & INESC TEC, Portugal

Daniel Diaz
Université Paris 1, France

Brett Drury
Universidade de São Paulo, Brazil

Jean-Marie Favre
Université Joseph Fourier, Grenoble, France

Luís Ferreira
Instituto Politécnico do Cávado e Ave, Portugal

Jean-Christophe Filliâtre
CNRS & Université Paris Sud, France

Niklas Fors
Lund University, Sweden

Pablo Gamallo
Universidade de Santiago de Compostela, Spain

Alda Lopes Gançarski
Institut Mines-Télécom/Télécom SudParis, France

Xavier Gómez Guinovart
Universidade de Vigo, Spain

Ulrich Heid
Universität Hildesheim, Germany

Pedro Rangel Henriques
Universidade do Minho, Portugal

Mirjana Ivanovic
University of Novi Sad, Serbia

Jan Janoušek
Czech Technical University in Prague, Czech Republic

Ján Kollár
Technical University of Košice, Slovakia

Tomaž Kosar
Univerza v Mariboru, Slovenia

3^{rd} Symposium on Languages, Applications and Technologies (SLATE'14).
Editors: Maria João Varanda Pereira, José Paulo Leal, and Alberto Simões
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Eugenijus Kurilovas
Vilnius Gediminas Technical University, Lithuania

José Paulo Leal
Universidade do Porto, Portugal

António Menezes Leitão
INESC-ID & Universidade de Lisboa, Portugal

Giovani Librelotto
Universidade Federal Santa Maria, Brazil

João Correia Lopes
Universidade do Porto & INESC TEC, Portugal

Ivan Lukovic
University of Novi Sad, Serbia

Paulo Matos
Instituto Politécnico de Bragança, Portugal

Marjan Mernik
Univerza v Mariboru, Slovenia

Michal Krátký
VŠB – Technical University of Ostrava, Czech Republic

Hugo Gonçalo Oliveira
Universidade de Coimbra, Portugal

Nuno Oliveira
Universidade do Minho, Portugal

Alexander Paar
TWT GmbH Science and Innovation, Germany

Lluís Padró
Universitat Politècnica de Catalunya, Spain

Thiago Pardo
Universidade de São Paulo, Brazil

Maria João Varanda Pereira
Instituto Politécnico de Bragança, Portugal

Jaroslav Porubän
Technical University of Košice, Slovakia

Ricardo Queirós
Instituto Politécnico do Porto, Portugal

José Carlos Ramalho
Universidade do Minho, Portugal

Sebastian Rahtz
University of Oxford, United Kingdom

Cristina Ribeiro
Universidade do Porto & INESC TEC, Portugal

Ricardo Rocha
Universidade do Porto, Portugal

Casiano Rodriguez-Leon
Universidad de La Laguna, Spain

Dietmar Seipel
Universität Würzburg, Germany

José Luis Sierra
Universidad Complutense de Madrid, Spain

Josep Silva
Universitat Politècnica de València, Spain

Alberto Simões
Universidade do Minho, Portugal

Boštjan Slivnik
Univerza v Ljubljani, Slovenia

Peter Sloep
Open Universiteit, Netherlands

Simão Melo de Sousa
Universidade da Beira Interior, Portugal

Ralf Steinberger
EC – Joint Research Centre, Italy

Kari Systä
Tampere University of Technology, Finland

António Teixeira
Universidade de Aveiro, Portugal

Jörg Tiedemann
Uppsala University, Sweden

Guido Wachsmuth
Delft University of Technology, Netherlands

Yorick Wilks
Florida Institute for Human and Machine Cognition, USA

Sub Reviewers

Mário Rodrigues
Universidade de Aveiro, Portugal

Paula Christina Figueira Cardoso
Universidade de São Paulo, Brazil

Marcos Garcia
Universidade de Santiago de Compostela, Spain

Liliana Ferreira
Fraunhofer AICOS, Portugal

Organization Committee

Maria João Varanda
Instituto Politécnico de Bragança, Portugal

José Paulo Leal
Universidade do Porto, Portugal

Alberto Simões
Universidade do Minho, Portugal

Pedro Henriques
Universidade do Minho, Portugal

Nuno Ramos Carvalho
Universidade do Minho, Portugal

José Eduardo Fernandes
Instituto Politécnico de Bragança, Portugal

Paulo Matos
Instituto Politécnico de Bragança, Portugal

Paulo Alves
Instituto Politécnico de Bragança, Portugal