

Vldia Pinheiro
Gustavo Henrique Paetzold

STIL 2017

**XI Brazilian Symposium in Information and
Human Language Technology and Collocated
Events**

Proceedings of the Conference

**October 2 to 5, 2017.
Uberlndia, Minas Gerais.**

©2017 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. Re-publication of material from this volume requires permission by the copyright owners.

Editors' addresses:

Universidade de Fortaleza
Av. Washington Soares, 1321, Edson Queiroz
Fortaleza, CE - Brasil, 60811-905
vladiacelia@unifor.br

University of Sheffield
Western Bank
Sheffield, United Kingdom, S10 2TN
g.h.paetzold@sheffield.ac.uk

XI Brazilian Symposium in Information and Human Language Technology

This volume contains the papers presented at the XI Symposium in Information and Human Language Technology (STIL 2017) and at the V Workshop on Portuguese Description (JDP), held on October 2-5, 2017 in Uberlândia, Brazil.

STIL is the bi-annual Language Technology event supported by the Brazilian Computer Society (SBC) and by the Brazilian Special Interest Group on Natural Language Processing (CE-PLN). The conference has a multidisciplinary nature and covers a broad spectrum of disciplines related to Human Language Technology, such as Linguistics, Computer Science, Psycholinguistics, Information Science, and others. It aims at bringing together both academic and industrial participants working on those areas.

The topics of interest centered around work in human language technology in general, such as Natural Language Resources & Tools, Corpus Linguistics, Text Classification, Sentiment Analysis and Opinion Mining, Information Extraction & Retrieval, Statistical and Machine Learning Methods, Natural language interfaces, Summarization, Terminology, Lexicology and Lexicography, to name a few.

We received 51 submissions from Brazil, Portugal, Spain, USA and Denmark. Each paper was reviewed by at least two members of the Program Committee, which had 48 members from 6 countries and 30 institutions. After a rigorous reviewing process, 14 papers were selected for oral presentation, and 6 papers were selected for poster presentation.

We thank the authors for their submissions, the program committee for their hard work, invited speakers, SBC staff and the Local and General Chairs of STIL 2017.

October 2017

Vlória Pinheiro
Gustavo Henrique Paetzold

Acknowledgments

The Program Committee chairs acknowledge the financial support to the conference provided by the Brazilian Computer Society (SBC), and the Federal University of Uberlândia (UFU). We thank the Program Committees of the XI Brazilian Symposium in Information and Human Language Technology and Collocated Events for the reviews that they produced. Last but not least, we are grateful to the local organization led by Prof. Gina Maira Barbosa de Oliveira (BRACIS 2017 General Chair) and Paulo Henrique Ribeiro Gabriel (STIL 2017 Local Chair).

October 2017

Vlória Pinheiro
Gustavo Henrique Paetzold

Program Chairs

Vladia Pinheiro (Universidade de Fortaleza - UNIFOR, Brazil)

Gustavo H. Paetzold (Universidade de Sheffield, UK)

Local Chair

Paulo Henrique Ribeiro Gabriel (Federal University of Uberlandia - UFU, Brazil)

Program Committee

Alberto Simoes (University of Minho – UMinho, Portugal)

Alexandre Rademaker (Fundaao Getulio Vargas – FGV & IBM Research, Brazil)

Aline Villavicencio (Federal University of Rio Grande do Sul – UFRGS, Brazil)

Andre Adami (University of Caxias do Sul – UCS, Brazil)

Antonio Branco (University of Lisbon – ULisboa, Portugal)

Arnaldo Candido Junior (Federal Technological University of Parana – UTFPR, Brazil)

Carlos A. Prolo (Federal University of Rio Grande do Norte – UFRN, Brazil)

Carlos Ramisch (Aix Marseille Universite – AMU, France)

Cassia Trojahn dos Santos (IRIT & University of Toulouse 2 – UTM2, France)

Christopher Shulby (University of Sao Paulo – USP/ICMC, Brazil)

Clarissa Xavier (Pontifical Catholic University of Rio Grande do Sul – PUCRS, Brazil)

Claudia de Freitas (Pontifical Catholic University of Rio de Janeiro – PUC-Rio, Brazil)

Daniel Lucredio (Federal University of Sao Carlos – UFSCar, Brazil)

Daniel Muller (Federal University of Rio Grande do Sul – UFRGS, Brazil)

Diana Santos (Linguatca & University of Oslo – UiO, Norway)

Eraldo Fernandes (Federal University of Mato Grosso do Sul – UFMS, Brazil)

Erick Maziero (University of Sao Paulo – USP/ICMC, Brazil)

Gustavo H. Paetzold (University of Sheffield, UK)

Helena Caseli (Federal University of Sao Carlos – UFSCar, Brazil)

Heliana Mello (Federal University of Minas Gerais – UFMG, Brazil)

Hugo Gonalo Oliveira (University of Coimbra – UC, Portugal)

Ivandre Paraboni (University of Sao Paulo – USP/EACH, Brazil)

Jorge Baptista (University of Algarve – UAAl, Portugal)

Leandro Mendonca de Oliveira (Brazilian Agricultural Research Corporation – EMBRAPA, Brazil)

Leonel Alencar (Federal University of Ceara - UFC, Brazil)

Livy Real (Universidade Federal do Parana, Brazil)

Lucelene Lopes (Pontifical Catholic University of Rio Grande do Sul – PUCRS, Brazil)

Luciano Barbosa (IBM Research, Brazil)

Marcelo Finger (University of Sao Paulo – USP/IME, Brazil)

Maria das Graas Nunes (University of Sao Paulo – USP/ICMC, Brazil)

Marcio Dias (Federal University of Goias – UFG, Brazil)

Mário Silva (INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Portugal)
Nelson Neto (Federal University of Pará - UFPA, Brazil)
Norton Roman (University of São Paulo – USP/EACH, Brazil)
Osvaldo de Oliveria Jr. (University of São Paulo, Brazil)
Paloma Moreda (University of Alicante – UA, Spain)
Pedro Balage Filho (University of São Paulo – USP/ICMC, Brazil)
Renata Vieira (Pontifical Catholic University of Rio Grande do Sul – PUCRS, Brazil)
Roseli Ap. Francalin Romero (University of São Paulo – USP/ICMC, Brazil)
Ruy Milidiú (Pontifical Catholic University of Rio de Janeiro – PUC-Rio, Brazil)
Sandra Aluísio (University of São Paulo – USP/ICMC, Brazil)
Sérgio de Freitas (University of Brasília – UnB, Brazil)
Stella Tagnin (University of São Paulo – USP/FFLCH, Brazil)
Thiago Pardo (University of São Paulo – USP/ICMC, Brazil)
Valéria Feltrim (State University of Maringá – UEM, Brazil)
Valeria de Paiva (Nuance Communications, USA)
Vera Lúcia Strube de Lima (Pontifical Catholic University of Rio Grande do Sul – PUCRS, Brazil)
Vlória Pinheiro (University of Fortaleza – UNIFOR, Brazil)

Natural Language Processing Steering Committee

Vlória Pinheiro (coordinator) (Universidade de Fortaleza - UNIFOR, Brazil)
Maria Cláudia de Freitas (Pontifícia Universidade Católica do Rio de Janeiro - PUC-Rio, Brazil)
Valéria Delisandra Feltrim (Universidade Estadual de Maringá - UEM, Brazil)
Carlos Augusto Prolo (Universidade Federal do Rio Grande do Norte - UFRN, Brazil)
Alexandre Rademaker (Fundação Getúlio Vargas - FGV & IBM Research, Brazil)

Contents

I	Conference Papers	10
1	Short Papers	11
	Uma Ferramenta para Identificar Desvios de Linguagem na Língua Portuguesa <i>Jonathan Nau and Aluizio Haendchen Filho and Guilherme Passero and Vinicius Cavaco</i>	12
	Estudo exploratório de categorias gramaticais com potencial de indicadores para a Análise de Sentimentos <i>Júlia Rodrigues and Adriana Pagano and Emerson Paraíso</i>	17
	Geração de perguntas e respostas para a base de conhecimento de um chatterbot educacional <i>Joyce Martins and Camila Martins</i>	22
	A study on irony within the context of 7x1-PT corpus <i>Silvia Moraes and Rackel Machado and Matheus Redecker and Rafael Cadaval and Felipe Meneguzzi</i>	27
	Improving Opinion Summarization by Assessing Sentence Importance in On-line Reviews <i>Rafael Anchiêta and Rogerio Figueredo Sousa and Raimundo Moura and Thiago Pardo</i>	32
	Wheel of Life: an initial investigation. Topic-Related Polarity Visualization in Personal Stories <i>Henrique Santos and Greice Pinho and Jackson Pinheiro and Renata Vieira</i>	37
2	Full Papers	42
	Investigating Opinion Mining through Language Varieties: a Case Study of Brazilian and European Portuguese tweets <i>Douglas Vitorio and Ellen Polliana Ramos Souza and Ingrid Pereira and Adriano Oliveira</i>	43
	A Rule-based Semantic Annotator: Adding top-level ontology Tags <i>Guidson Coelho Andrade and Alcione Oliveira and Alexandra Moreira</i>	53

CONTENTS

Processo de construção de um corpus anotado com Entidades Geológicas visando REN <i>Daniela Amaral and Sandra Collovini and Anny Figueira and Renata Vieira and Renata Vieira and Marco Gonzalez</i>	63
Formação de gentílicos a partir de topônimos: descrição linguística e aprendizado automático <i>Roger Alfredo Marci Rodrigues Antunes and Thiago Pardo and Gladis Maria Barcelos Almeida</i>	73
Visualização de glossário em sistemas de recuperação de informação <i>Glauber Vaz and Leandro Mendonça Oliveira and Ivo Pierozzi Jr.</i>	83
Avaliando a similaridade semântica entre frases curtas através de uma abordagem híbrida <i>Allan Silva and Sandro Rigo and Isa Mara Alves and Jorge Barbosa</i>	93
A Comparative Study for Sentiment Analysis on Election Brazilian News <i>Caio Magno Aguiar Carvalho and Hitoshi Nagano and Allan Kardec Barros</i> . . .	103
Influência de Técnicas Não-supervisionadas de Redução de Dimensionalidade para Organização Flexível de Documentos <i>Beatriz Lima and Fernanda Eustáquio and Tatiane Nogueira</i>	112
Portuguese Word Embeddings: Evaluating on Word Analogies and Natural Language Tasks <i>Nathan Hartmann and Erick Fonseca and Christopher Shulby and Marcos Treviso and Jéssica Silva and Sandra Aluísio</i>	122
Utilizando Features Linguísticas Genéricas para Classificação de Triplas Relacionais em Português <i>George Barbosa and Daniela Barreiro Claro</i>	132
Normalizador de Texto para Língua Portuguesa baseado em Modelo de Linguagem <i>Patrick Bard and Renan Lopes Luis and Silvia Moraes</i>	142
Evaluating Word Embeddings for Sentence Boundary Detection in Speech Transcripts <i>Marcos Treviso and Christopher Shulby and Sandra Aluísio</i>	151
Análise de Medidas de Similaridade Semântica na Tarefa de Reconhecimento de Implicação Textual <i>David Feitosa and Vladia Pinheiro</i>	161
FrameFOR – Uma Base de Conhecimento de Frames Semânticos para Perícias de Informática <i>Ravi Barreira and Vladia Pinheiro and Vasco Furtado</i>	171
II V Jornada de Descrição do Português	181
3 Papers	184
A Implementação de uma Minigramática do Português Brasileiro sob a Perspectiva da LFG <i>Daniel Soares and Francisco Nogueira and Leonel Figueiredo Alencar</i>	185

CONTENTS

Constituição de Um Dicionário Eletrônico Trilíngue Fundado em Frames a partir da Extração Automática de Candidatos a Termos do Domínio do Turismo <i>Simone Rodrigues Peron-Corrêa and Tiago Timponi Torrent</i>	193
A Modelagem Computacional do Domínio dos Esportes na FrameNet Brasil <i>Alexandre Diniz Costa and Tiago Timponi Torrent</i>	201
Descrição e modelagem de construções interrogativas QU- em Português Brasileiro para o desenvolvimento de um chatbot <i>Natália Duarte Marção and Tiago Timponi Torrent and Ely Edison Silva Matos</i> .	209
Construções de Estrutura Argumental no âmbito do Constructicon da FrameNet Brasil: proposta de uma modelagem linguístico-computacional <i>Vânia Gomes Almeida and Tiago Timponi Torrent</i>	217
Investigação Preliminar Sobre a Prosódia Semântica de Verbos de Elocução: o Caso do Verbo "Confessar" <i>Barbara C. Ramos</i>	224
Uma Proposta Metodológica para a Categorização Automatizada de Atrações Turísticas a partir de Comentários de Usuários em Plataformas Online <i>Vanessa Maria Ramos Lopes Paiva and Tiago Timponi Torrent</i>	232
Sofrer uma ofensa, Receber uma advertência: Verbos-suporte Conversos de 'Fazer' no Português do Brasil <i>Claúdia D. Barros and Nathalia P. Calcia and Oto A. Vale</i>	240
Os Provérbios em manuais de ensino de Português Língua Não Materna <i>Sónia Reis and Jorge Baptista</i>	247
Prosody, syntax, and pragmatics: insubordination in spoken Brazilian Portuguese <i>Giulia Bossaglia and Heliana Mello and Tommaso Raso</i>	256
As bases de dados verbais ADESSE e ViPER: uma análise contrastiva das construções locativas em espanhol e em português <i>Roana Rodrigues and Oto Vale and Laura Alonso Alemany</i>	266