

IX Brazilian Symposium in Information and Human Language Technology

(STIL 2013)

**Fortaleza, Ceara, Brazil
21-23 October 2013**

ISBN: 978-1-62993-548-5

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2013) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2014)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

TABLE OF CONTENTS

| | |
|--|-----|
| AgreeCalc: Uma Ferramenta para Análise da Concordância entre Múltiplos Anotadores (AgreeCalc: A Tool for the Analysis of Agreement Between Multiple Annotators) | 1 |
| <i>Alexandre Rossi Alvares, Norton Trevisan Roman</i> | |
| MetaAnn: Um Gerador de Ferramentas para Anotação de Textos (MetaAnn: A Generator of Text Annotation Tools) | 11 |
| <i>Tiago Emanuel Infante Missão, Norton Trevisan Roman</i> | |
| Improving CoGrOO: The Brazilian Portuguese Grammar Checker | 21 |
| <i>William D. Colen M. Silva, Marcelo Finger</i> | |
| Uma Investigação sobre Algoritmos de Diferentes Abordagens de Aprendizado Supervisionado na Classificação de Papéis Retóricos em Resumos Científicos (Investigating Algorithms from Different Approaches of Supervised Learning for the Classification of Rhetorical Roles in Scientific Abstracts) | 30 |
| <i>Vinicius M. A. de Souza, Valéria D. Feltrim</i> | |
| O Subframe Sentença no Complexo Frame Processo de Conhecimento no Direito Processual Civil (The Sentencing Subframe in the Complex Knowledge Process Frame in the Civil Procedural Law) | 40 |
| <i>Thaís Domênica Minghelli, Anderson Bertoldi, Rove Chishman</i> | |
| Subtopic Annotation in a Corpus of News Texts: Steps Towards Automatic Subtopic Segmentation | 49 |
| <i>Paula C. F. Cardoso, Maite Taboada, Thiago A. S. Pardo</i> | |
| O Reconhecimento de Entidades Nomeadas por meio de Conditional Random Fields para a Língua Portuguesa (Named Entity Recognition with Conditional Random Fields for the Portuguese Language) | 59 |
| <i>Daniela O. F. do Amaral, Renata Vieira</i> | |
| Análise Automática de Coerência Usando o Modelo Grade de Entidades para o Português (Automatic Coherence Analysis Using the Entity-Grid Model for Portuguese) | 69 |
| <i>Alison R. P. Freitas, Valéria D. Feltrim</i> | |
| Aplicando Pontos de Corte para Listas de Termos Extraídos (Applying Cut-off Points to Lists of Extracted Terms) | 79 |
| <i>Lucelene Lopes, Renata Vieira</i> | |
| Geração de Expressões de Referência usando Relações Espaciais (Referring Expression Generation Using Spatial Relations) | 88 |
| <i>Diego dos Santos Silva, Ivandré Paraboni</i> | |
| Mac-Morpho Revisited: Towards Robust Part-of-Speech Tagging | 98 |
| <i>Erick Rocha Fonseca, João Luís G. Rosa</i> | |
| Implementação de um Separador Silábico Gratuito Baseado em Regras Linguísticas para o Português Brasileiro (The Implementation of a Free Syllabification Algorithm Based on Linguistic Rules for Brazilian Portuguese) | 108 |
| <i>Willian Rocha, Nelson Neto</i> | |
| Text Simplification as Tree Transduction | 116 |
| <i>Gustavo H. Paetzold, Lucia Specia</i> | |
| Automatic Disambiguation of Homographic Heterophone Pairs Containing Open and Closed Mid Vowels | 126 |
| <i>Christopher Shulby, Gustavo Mendonça, Vanessa Marquifável</i> | |
| Desambiguação Lexical de Sentido com uso de Informação Multidocumento por meio de Redes de Co-ocorrência (Word Sense Disambiguation with the Use of Multi-document Information with Co-occurrence Nets) | 138 |
| <i>Fernando Antônio Asevedo Nóbrega, Thiago Alexandre Salgueiro Pardo</i> | |
| Realização Superficial baseada em Regras (Rule-based Surface Realisation) | 148 |
| <i>Douglas F. P. da Silva Junior, Eder M. de Novais, Ivandré Paraboni</i> | |
| JWN-Br - Uma API Java para a WordNet.Br (JWN-Br - An Java API for Wordnet.Br) | 153 |
| <i>Vitor Machado Oliveira, Norton Trevisan Roman</i> | |
| Geração de features para resolução de correferência: Pessoa, Local e Organização (Feature Generation for Coreference Resolution: Person, Location and Organization) | 158 |
| <i>Evandro B. Fonseca, Renata Vieira, Aline A. Vanin</i> | |
| Modalidade no Português Brasileiro: as estruturas condicionais (Modality in Brazilian Portuguese: The Conditional Structures) | 163 |
| <i>Priscila Osório Côrtes, Heliana Mello</i> | |

| | |
|--|-----|
| Um repositório de verbos para a anotação de papéis semânticos disponível na web (A Verb Repository for Semantic Role Labeling Available in the Web) | 168 |
| <i>Magali Sanches Duran, Jhonata Pereira Martins, Sandra Maria Aluísio</i> | |
| Entity-centric Sentiment Analysis on Twitter data for the Portuguese Language | 173 |
| <i>Marlo Souza, Renata Vieira</i> | |
| Approaches for Helping Brazilian Students Improve their Scientific Writings | 178 |
| <i>Ethel Schuster, Rick Lizotte, Sandra M. Aluísio, Carmen Dayrell</i> | |
| BrGram: uma gramática computacional de um fragmento do português brasileiro no formalismo da LFG (BrGram: A Computational Grammar for a Fragment of Brazilian Portuguese in the LFG Formalism) | 183 |
| <i>Leonel F. de Alencar</i> | |
| Geração de instruções em mundos virtuais: primeiros passos (Generation of Instructions in Virtual Worlds: First Steps) | 189 |
| <i>Diego Cardozo Sandrim, Felipe Freire, Ivandré Paraboni</i> | |
| Extração de Vocabulário Multilíngue para Tradução em Domínios Especializados (Multilingual Vocabulary Extraction for Machine Translation in Specialized Domains) | 194 |
| <i>Lucas Welter Hilgert, Renata Vieira</i> | |
| Corpus-amostra Coelho Netto: compilação, anotação e ocorrências em textos literários dos séc. XIX e XX (Corpus-sample Coelho Netto: Compilation, Annotation and Occurrences in Literary Texts from XIX and XX Centuries) | 199 |
| <i>Francimary Macêdo Martins</i> | |
| Acoustic, Phonetic and Prosodic Features of Parkinson's Disease Speech | 205 |
| <i>Jorge Proença, Arlindo Veiga, Sara Candeias, Fernando Perdigão</i> | |
| Classificação de Prioridade de Tweets utilizando Máquinas de Vetor de Suporte (Tweet Priority Classification Using Support Vector Machines) | 210 |
| <i>Vinícius Pazzini, Tiago Schenkel, Mikael Poetsch, Ricardo Matsumura Araujo</i> | |
| An Evaluation of the Brazilian Portuguese LIWC Dictionary for Sentiment Analysis | 215 |
| <i>Pedro P. Balage Filho, Thiago Alexandre Salgueiro Pardo, Sandra M. Aluísio</i> | |
| Uma investigação no apoio da avaliação da usabilidade em Sistemas Sociais usando Processamento da Linguagem Natural (An Investigation on the Support of the Usability Evaluation in Social Systems Using Natural Language Processing) | 220 |
| <i>Marília S. Mendes, Elizabeth Furtado, Miguel Franklin</i> | |
| Uma Ferramenta para Projeto de Sistemas de Diálogos para Call Center Baseados em Asterisk (A Tool for the Design of Dialog Systems for Asterisk-based Call Centers) | 225 |
| <i>Estillac Borges Filho, Pedro Batista, Aldebaro Klautau</i> | |
| Author Index | |