Volume 2, Número 1 – Abril 2010

LinguaMÁTICA

ISSN: 1647–0818

Conteúdo

Editorial

Este é o terceiro número da ***Linguamática*** *e o primeiro de 2010, um número que termina o percurso da revista ao longo de um ano. Trata-se de uma edição especial com artigos seleccionados do* Sétimo Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana *(STIL'09), o que demonstra o interesse da nossa comunidade científica na* ***Linguamática****.*

Todos os artigos deste número especial são publicados na seccção dedicada aos Artigos de Investigação*. Agradecemos a colaboração dos autores seleccionados e dos organizadores do STIL na elaboração deste número da* ***Linguamática****.*

Finalmente, queremos marcar mais uma etapa na revista celebrando a indexação da ***Linguamática*** *em catálogos de bibliotecas digitais e em índices públicos de revistas electrónicas, entre os quais salientamos o* Latindex — Sistema Regional de información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal —*, o* DOAJ — Directory of Open Access Journals —*, o* Google Scholar *e o* The Linguist List*.*

Xavier Gómez Guinovart
José João Almeida
Alberto Simões

Prólogo

Uma visão geral dos avanços no Simpósio de Tecnologia da Informação e Linguagem Humana

Esta edição especial da Linguamática contém uma seleção dos artigos apresentados no 7º Simpósio de Tecnologia da Informação e Linguagem Humana (STIL 2009), que ocorreu de 8 a 11 de setembro de 2009 na Universidade de São Paulo (campus São Carlos), Brasil (`http://www.nilc.icmc.usp.br/til/stil2009_English`*). O STIL*[1] *é o evento anual de Tecnologia da Linguagem apoiado pela Sociedade Brasileira de Computação (SBC) e pela Comissão Especial de Processamento de Linguagem Natural. Este evento tem um caráter multidisciplinar, abrangendo um amplo espectro de disciplinas relacionadas à Tecnologia da Linguagem Humana, tais como Lingüística, Ciências da Computação, Psicologia, Ciência da Informação, entre outros, e tem por objetivo reunir participantes acadêmicos e da indústria que atuam nessas áreas.*

Os tópicos de interesse anunciados no Call for Papers estiveram centrados em torno dos trabalhos em tecnologia da linguagem humana em geral realizados a partir de perspectivas tão diversas como Ciências da Computação, Lingüística e Ciência da Informação, incluindo entre outros a mineração de texto, processamento da linguagem escrita e falada, a terminologia, lexicologia e lexicografia, modelagem e gestão de conhecimento e geração de linguagem natural. Foram submetidos 60 artigos longos e 26 curtos. Cada proposta foi analisada por três membros do Comitê de Programa, composto por 88 pesquisadores de 13 países e 45 instituições.

Após um rigoroso processo de revisão 18 artigos completos e 12 curtos foram selecionados, com taxas de aceitação de 30% e 42%, respectivamente. Os autores dos artigos completos foram convidados a submeter versões estendidas e revisadas dos seus trabalhos para esta edição especial, passando por um novo processo de revisão, desta vez pelos revisores da Linguamática, que selecionaram 7 dos artigos submetidos.

Estes artigos representam uma amostra do rico e variado trabalho apresentado no STIL e envolvem pesquisadores de instituições acadêmicas e industriais no Brasil, Portugal e França. Por exemplo, o primeiro artigo, Identificação de expressões multipalavra em domínios específicos *de Aline Villavicencio et al., propõe uma abordagem para a identificação de Expressões Multipalavra, tais como compostos nominais e verbos frasais, em corpora técnicos. A proposta apresentada combina medidas de associação com informações linguísticas e de alinhamentos lexicais, e o artigo examina a influência de diversos fatores sobre o seu desempenho.*

Os dois próximos artigos são relacionados a aplicações de PLN. Em Classificação

[1] Este evento era anteriormente conhecido como TIL (Workshop de Informação e Tecnologia da Linguagem Humana).

automática de textos por período literário utilizando compressão de dados através do PPM-C, *Bruno Barufaldi et al. propõem a aplicação do método Prediction by Partial Matching (PPM) para a tarefa de classificação de textos de acordo com períodos literários da literatura brasileira. Já Carolina Scarton e Sandra Aluísio, em* Análise da inteligibilidade de textos via ferramentas de processamento de língua natural: adaptando as métricas do Coh-Metrix para o Português*, investigam a adaptação de métricas da ferramenta Coh-Metrix para o português do Brasil (Coh-Metrix-Port), primeiramente avaliando as diferenças entre textos complexos para adultos e versões mais simples para crianças e também analisando o desempenho de classificadores para discriminar textos dedicados a adultos e a crianças, que podem ser usados para avaliar a simplicidade de textos disponíveis na Web.*

O quarto artigo Caracterização e processamento de expressões temporais em português *de Caroline Hagège, Jorge Baptista e Nuno Mamede também aborda a questão do tratamento de expressões, mas desta vez o foco é em expressões temporais tais como* de manhã *e* nesta semana*. Os autores propõem uma classificação para estas expressões do português e apresentam uma ferramenta de anotação delas em corpora.*

Quanto a construção de recursos linguísticos para o português, o artigo Extração de relações semânticas entre palavras a partir de um dicionário: o PAPEL e sua avaliação *de Hugo Oliveira, Diana Santos e Paulo Gomes apresenta o PAPEL, um recurso lexical que contém relações entre palavras, como sinonímia, automaticamente extraídas de um dicionário através de regras, discutindo ainda uma avaliação do mesmo.*

Outra tarefa abordada neste volume é a de sumarização, no artigo Estratégias de seleção de conteúdo com base na CST (Cross-document Structure Theory) para sumarização automática multidocumento *de Maria Jorge e Thiago Pardo. Os autores discutem a definição, formalização e avaliação de estratégias de seleção de conteúdo para sumarização automática multidocumento com base na teoria discursiva Cross-document Structure Theory.*

Por fim a tarefa de entendimento e linguagem natural é abordada no artigo Um analisador semântico inferencialista de sentenças em linguagem natural *de Vladia Pinheiro et al, onde é descrito o Analisador Semântico Inferencialista (SIA), um raciocinador semântico sobre o conteúdo inferencial de conceitos e padrões de sentenças, avaliado em um sistema de extração de informações sobre crimes.*

Nossos agradecimentos para os editores da Linguamática e revisores dos artigos tanto da Linguamática quanto do STIL 2009.

Aline Villavicencio
Horácio Saggion
Maria das Graças Volpe Nunes
Thiago Pardo

Comissão Científica

Alberto Álvarez Lugrís, Universidade de Vigo
Alberto Simões, Universidade do Minho
Aline Villavicencio, Universidade Federal do Rio Grande do Sul
Álvaro Iriarte Sanroman, Universidade do Minho
Ana Frankenberg-Garcia, ISLA e Universidade Nova de Lisboa
Anselmo Peñas, Universidad Nacional de Educación a Distancia
Antón Santamarina, Universidade de Santiago de Compostela
António Teixeira, Universidade de Aveiro
Belinda Maia, Universidade do Porto
Carmen García Mateo, Universidade de Vigo
Diana Santos, SINTEF ICT
Ferran Pla, Universitat Politècnica de València
Gael Harry Dias, Universidade Beira Interior
Gerardo Sierra, Universidad Nacional Autónoma de México
German Rigau, Euskal Herriko Unibertsitatea
Helena de Medeiros Caseli, Universidade Federal de São Carlos
Horacio Saggion, University of Sheffield
Iñaki Alegria, Euskal Herriko Unibertsitatea
Joaquim Llisterri, Universitat Autònoma de Barcelona
José Carlos Medeiros, Porto Editora
José João Almeida, Universidade do Minho
José Paulo Leal, Universidade do Porto
Joseba Abaitua, Universidad de Deusto
Lluís Padró, Universitat Politècnica de Catalunya
Maria Antònia Martí Antonín, Universitat de Barcelona
Maria das Graças Volpe Nunes, Universidade de São Paulo
Mercè Lorente Casafont, Universitat Pompeu Fabra
Mikel Forcada, Universitat d'Alacant
Nieves R. Brisaboa, Universidade da Coruña
Pablo Gamallo Otero, Universidade de Santiago de Compostela
Salvador Climent Roca, Universitat Oberta de Catalunya
Susana Afonso Cavadas, University of Sheffield
Tony Berber Sardinha, Pontifícia Universidade Católica de São Paulo
Xavier Gómez Guinovart, Universidade de Vigo