

Volume 4, Número 1 – Abril 2012

LinguaMÁTICA

ISSN: 1647-0818

Editores Convidados

Diana Santos

Cristina Mota

Cláudia Freitas

Luís Costa

Editores

Alberto Simões

José João Almeida

Xavier Gómez Guinovart

Conteúdo

Porquê o Págico? Razões para uma avaliação conjunta <i>Diana Santos</i>	1
A lusofonia na Wikipédia em 150 tópicos <i>Cláudia Freitas</i>	9
Tirando o chapéu à Wikipédia: A coleção do Págico e o Cartola <i>Alberto Simões, Luís Costa e Cristina Mota</i>	19
Uma abordagem ao Págico baseada no processamento e análise de sintagmas dos tópicos <i>Ricardo Rodrigues, Hugo Gonçalo Oliveira e Paulo Gomes</i>	31
Medindo o precipício semântico <i>Nuno Cardoso</i>	41
O desafio da participação humana do IT-Coimbra no Págico <i>A. Veiga, C. Lopes, D. Celorico, J. Proença, F. Perdigão e S. Candeias</i>	49
Do tópico às respostas: do processo humano à sua simulação <i>Luísa Coheur e Ângela Costa</i>	53
Desafios na recolha de informação baseada na Wikipédia portuguesa com o Págico <i>João Miranda</i>	61
O que é uma resposta? Notas de uns avaliadores estafados <i>Cláudia Freitas, Paulo Rocha, Cristina Mota, Luís Costa e Diana Santos</i>	67
Resultados págicos: participação, medidas e pontuação <i>Cristina Mota</i>	77
Balanço do Págico e perspetivas de futuro <i>Diana Santos, Cristina Mota, Alberto Simões, Luís Costa e Cláudia Freitas</i>	93

Editorial

É com grande orgulho que observámos o início do quarto ano da Linguamática. Não é fácil garantir uma publicação periódica, quanto mais quando é dedicada a conteúdo científico, e quando somos forçados a publicar sempre nos mesmos sítios bem cotados. Mas só podemos ser uma publicação de referência se conseguirmos publicar, e por isso queremos agradecer a todos os autores que já participaram nesta aventura, e a todos os membros da comissão científica que sempre nos ajudaram a definir o melhor conteúdo.

Este quarto ano começa com a segunda edição especial da Linguamática que inaugura um novo visual. Sabemos que o conteúdo da revista é o importante, mas se nos der gozo lê-la, o conteúdo torna-se ainda mais cativante. Esperamos que gostem do novo estilo visual.

*Esta edição especial resume uma avaliação conjunta que visa a procura de informação, e nomeadamente a procura de informação na Wikipédia Portuguesa – Páxico (**P**ortuguês **M**ágico).*

As avaliações (conjuntas) são actividades que nos merecem o maior respeito. Mas porquê o nosso interesse em avaliações conjuntas? Por experiência própria sabemos que para além do enorme esforço ligado ao trabalho de organização, a “simples” participação traz inevitavelmente a certeza da dúvida: Por um lado é o volume de dados que torna o possível, impossível; o calculável, impraticável; e o trivial, incerto! Por outro lado é aquele algoritmo brilhante que acaba por não ser realista, ou aquele resultado tão esperado que não aparece.

Depois de ultrapassada essa fase traumática, verificamos que o nosso modelo até faz sentido; pode não ser o melhor, mais eficiente, e com melhores resultados, mas interessa e surpreende os outros participantes; e com um bocado de trabalho futuro, os tais avanços esperados até serão possíveis, calculáveis e pelo menos sonháveis! E concluímos que será interessantíssimo voltar a testar uma nova versão numa nova avaliação conjunta...

Estamos perante um laboratório de experimentação — e nisso acreditamos.

Xavier Gómez Guinovart

José João Almeida

Alberto Simões

Introdução à Edição Especial

Págico — Português Mágico

Esta edição pretende apresentar a uma audiência internacional o Págico, uma avaliação conjunta inovadora organizada pela Linguateca. O Págico foi organizado de forma ao seu encontro final ser um encontro satélite do PROPOR 2012, a cuja organização agradecemos o convite.

Dada a pouca vitalidade do mercado editorial português, por um lado, que não consegue publicar contribuições de interesse científico com pouca audiência, e a existência e pujança de uma revista de acesso aberto dedicada aos temas do processamento das línguas ibéricas, incluindo, naturalmente, o português, por outro, a Linguamática configurou-se como a opção de excelência.

Uma avaliação conjunta pressupõe sempre muito trabalho de organização e muitas decisões e escolhas que, para mais tarde serem úteis à comunidade que pretende servir, têm de ser documentadas e fundamentadas.

Por outro lado, é importante compreender a motivação e as estratégias dos participantes, e uma edição sobre uma avaliação conjunta tem como um dos pilares fundamentais a participação. Dar voz e disseminação aos participantes é, pois, uma das obrigações mais gratas dos organizadores, que ao divulgarem o trabalho feito com base na sua organização podem apreciar os frutos desse trabalho.

*Além disso, o resultado **conjunto** da organização e da participação permite construir recursos de treino e de avaliação, que, embora constituam um trabalho hercúleo para a organização, são talvez o que justifica em maior grau a existência de um volume de uma revista dedicada a essa iniciativa, porque construiu algo que ficará – esperamos que por muito tempo – disponível a tantos quantos queiram*

- *investigar a recolha de informação em português*
- *estudar a cultura lusófona*
- *criar sistemas de resposta automática a perguntas, de recolha de informação (RI), de visualização de resultados, etc.*
- *estudar a wikipédia e a sua evolução.*

A presente edição segue o seguinte formato (em que todas as contribuições foram revistas por dois elementos da comissão científica):

- 1. vários artigos, escritos por membros da organização do Págico, sobre as opções tomadas e a motivação das mesmas, que podemos considerar pré-avaliação;*
- 2. os artigos dos participantes;*
- 3. vários artigos, outra vez da organização, descrevendo os resultados do Págico: os problemas e dificuldade da avaliação, a pontuação dos participantes, os recursos finalmente produzidos, e um balanço crítico da iniciativa.*

Embora com esta estrutura tenhamos tentado seguir a evolução temporal da própria avaliação conjunta, e tentássemos que o volume pudesse ser lido do princípio ao fim como um livro, os artigos não foram escritos em sequência. Pelo contrário, encorajámos que fossem mencionadas questões e dados apenas encontrados em artigos “posteriores”, quando tal fizesse sentido.

Esperamos que esta edição possa contribuir para um maior conhecimento não só da iniciativa em si mas também dos problemas associados ao processamento da wikipédia em português, à compreensão de utilizadores humanos, e à criação de consultas em RI.

*Diana Santos
Cristina Mota
Cláudia Freitas
Luís Costa*

Comissão Científica Convidada

Alberto Simões, Universidade do Minho

António Teixeira, Universidade de Aveiro

Belinda Maia, Universidade do Porto

Cláudia Freitas, Pontifícia Universidade Católica do Rio de Janeiro e FCCN/Linguatca

Cristina Mota, FCCN/Linguatca

Diana Santos, Universidade de Oslo e FCCN/Linguatca

Fernando Perdigão, Universidade de Coimbra

José João Dias de Almeida, Universidade do Minho

Luís Costa, FCCN/Linguatca

Luísa Coheur, INESC-ID e Instituto Superior Técnico

Paulo Gomes, Universidade de Coimbra

Sandra Aluísio, Universidade de São Paulo

Stella Tagnin, Universidade de São Paulo

Xavier Gómez Guinovart, Universidade de Vigo