

Índice

Prefácio	i
1 Utilização da programação declarativa para processamento do CETEMPúblico - <i>Agostinho Monteiro, Júlio Barbas e Nuno C. Marques</i>	1
1.1 Descrição do formato TXT/2	3
1.2 Exemplos de aplicação e resultados	4
2 Extracção de recursos de tradução - <i>Alberto Simões</i>	7
3 Novas ferramentas e recursos linguísticos para a tradução automática: por ocasião d’o fim do início de uma nova era no processamento da língua portuguesa - <i>Anabela Barreiro</i>	13
3.1 Tradução automática com conhecimento linguístico parafrástico	15
3.2 ReEscreve: um parafraseador monolíngue	15
3.3 ParaMT: um parafraseador bilingue/multilíngue	17
3.4 Recursos e metodologia adoptados na concepção dos parafraseadores	18
3.5 Avaliação quantitativa: primeiros resultados	21
3.6 Considerações finais	22
4 O corpus CONDIV e o estudo da convergência e divergência entre variedades do português - <i>Augusto Soares da Silva</i>	25
5 Os recursos da Linguateca ao serviço do desenvolvimento da tecnologia de fala na Microsoft - <i>Daniela Braga e Miguel Sales Dias</i>	29
5.1 A experiência da indústria: o impacto da Linguateca no desenvolvimento de produtos na Microsoft	31

5.2	Conclusão	32
6	Um estudo no COMPARA: a semântica dos compostos nominais - <i>Lílian Figueiró Teixeira e Rove Luiza de Oliveira Chishman</i>	35
6.1	Extração dos dados do COMPARA	37
6.2	Análise das relações semânticas	38
6.3	Considerações finais	40
7	Linguatca e Processamento de Linguagem Natural na Área da Saúde: Alguns Comentários e Sugestões - <i>Liliana Ferreira, António Teixeira e João Paulo da Silva Cunha</i>	43
7.1	MedAlert	44
7.2	Participação no Segundo HAREM	47
7.3	Conclusões e sugestões finais	48
8	Criação e expansão de geo-ontologias, dimensionamento de informação geográfica e reconhecimento de locais e seus relacionamentos em textos - <i>Marcirio Chaves</i>	49
8.1	Geographic Knowledge Base - GKB	51
8.2	Aplicações que utilizam as geo-ontologias geradas a partir da GKB	52
8.2.1	Sistemas de REM	52
8.2.2	Módulos de um sistema de recolha de informação geográfica	52
8.2.3	Interface de Motor de Pesquisa Geográfica	53
8.2.4	Interface para consultas a almanaques geo-temporais	53
8.3	Geograficidade de textos	56
8.4	Sistema de Extração, Anotação e Integração de conhecimento Geográfico - SEI-Geo	57
8.4.1	Avaliação do SEI-Geo	57
8.5	Considerações Finais	59
9	Relato sobre a parceria Linguatca-NILC - <i>Maria das Graças Volpe Nunes</i>	61
10	Uma abordagem estatística para a identificação de colocações verbais usando o projeto AC/DC em www.linguatca.pt - <i>Milena Uzeda Garrão e Maria Carmelita Padua Dias</i>	65

	137
10.1 Metodologia	66
10.1.1 O corpus utilizado: CETENFolha	66
10.1.2 Aplicação do filtro para padrões V+SN aos verbos mais frequentes	67
10.1.3 A aplicação do logaritmo de verossimilhança aos padrões V+(det)+N encabeçados pelos verbos mais frequentes no corpus .	67
10.1.4 Resultados e Edição Humana	68
10.2 Conclusões e trabalhos futuros	68
11 Novos rumos para a recuperação de informação geográfica em português - <i>Nuno Cardoso</i>	71
11.1 Compreendendo as consultas dos utilizadores	73
11.1.1 Reformulação automática de consultas	73
11.1.2 Consultas de âmbito geográfico	74
11.2 Rede de conhecimento	75
11.2.1 Fontes de informação	76
11.2.2 Características das fontes de informação	77
11.3 Trabalho desenvolvido até ao momento	79
11.3.1 QuerCol	81
11.3.2 REMBRANDT	82
11.3.3 MG4J	83
11.3.4 RENOIR	84
11.4 Avaliação do desempenho dos sistemas	84
12 Uso de marcadores estilísticos para a busca na Internet em português - <i>Rachel Aires</i>	87
13 Listas de frequência de palavras como marcadores de estilo no reconheci- mento de autoria - <i>Rui Sousa Silva</i>	93
14 Conversor de grafemas para fones baseado em regras para português - <i>Sara Candeias e Fernando Perdigão</i>	99
14.1 Arquitectura do sistema de conversão GR2PH	101
14.1.1 Subsistema de divisão silábica	101
14.1.2 Subsistema de marcação de tonicidade	102
14.1.3 Subsistema de transcrição para fones	102

14.2 Conclusão e trabalho futuro	104
15 Dez anos de convivência: um apanhado geral quanto ao uso dos recursos da Linguateca no Programa de Pós-Graduação em Ciência da Computação da PUCRS - Brasil - <i>Vera Lúcia Strube de Lima</i>	105
16 10 anos de Linguateca - depoimento - <i>Violeta Quental</i>	109
Referências	115
Índice	135