

ÍNDICE

Prefácio

Lista dos capítulos

Lista alfabética de autores

Capítulo 1. Avaliação conjunta

Diana Santos

1. Apresentação
2. O modelo da avaliação conjunta
 - 2.1 Modelos de avaliação anteriores
 - 2.2 Características principais
 - 2.3 Vantagens para os sistemas participantes
 - 2.4 Algumas críticas e limitações
3. Um pouco de história a nível internacional
 - 3.1 Contrapondo o TREC e o MUC
 - 3.2 O modelo do francês
 - 3.3 O AvalON, para o português
4. A implementação de uma actividade de avaliação conjunta
 - 4.1 Cartografia do problema
 - 4.2 A definição da tarefa
 - 4.3 Recursos de avaliação
 - 4.4 Medidas
 - 4.5 Alguns comentários provenientes da experiência prática
5. Avaliação de uma avaliação conjunta
6. Uma instanciação do modelo para o português

Capítulo 2. Organização e resultados morfolímpicos

Luís Costa, Paulo Rocha e Diana Santos

1. Apresentação
2. Historial
3. Lista dourada
 - 3.1 Directivas de classificação para cada forma
 - 3.2 Metodologia da compilação das várias formas
 - 3.3 Caracterização da lista dourada utilizada nas Morfolimpíadas
4. Criação dos textos
 - 4.1 Primeira fase: vários castores
 - 4.2 Segunda fase: castor-chefe
 - 4.3 Preparação dos ficheiros que foram distribuídos
 - 4.4 Caracterização dos textos
5. Processamento da saída dos sistemas
6. Resultados
 - 6.1 Critérios de comparação
7. Recurso
 - 7.1 Atomização
8. Três facetas das Morfolimpíadas
 - 8.1 Os sistemas como verificadores

8.2 Comparação com as Morpholympics

9. Comentários finais

Capítulo 3. Segmentação e análise morfológica do português com recursos léxicos limitados

Caroline Hagège

1. Apresentação geral do Smorph

2. O Smorph-PT: dados linguísticos e sua organização

2.1 Declaração de traços morfológicos

2.2 Definição de caracteres ASCII

2.3 Terminações distintas

2.4 Modelos flexionais

2.5 Entradas léxicas

2.6 Afixos

3. Autômato lexical

4. Exemplos de saída do Smorph

5. Algumas opções tomadas quanto ao desenvolvimento dos dados para o Smorph

6. A experiência das Morfolimpíadas

7. Conclusão e futuro do sistema

Capítulo 4. Dos processos de individuação e de categorização lexical

Ronaldo Teixeira Martins e Maria das Graças Volpe Nunes

1. Da individuação e da categorização lexical

2. Da correção ortográfica automática e da revisão gramatical de estilo

3. O projeto ReGra

4. Do léxico do ReGra

5. Do reconhecimento e da análise lexical no âmbito do ReGra

6. Da participação do ReGra nas Morfolimpíadas

Capítulo 5. Morfologia com sintaxe debaixo de olho

Eckhard Bick

1. Introdução

2. O PALAVRAS e o Palmorf-Avalon

2.1 A anatomia do Palmorf e sua integração no analisador sintático

2.2 Sequência de programas usados nas Morfolimpíadas

3. Alguns problemas na transformação de um desambiguador morfológico num analisador morfológico tradicional

3.1 Atomização

3.2 Categoria gramatical morfológica versus sintática

3.3 Etiquetas de não-especificação ou «portmanteau»

3.4 Rotinas de normalização de texto

4. Derivação

4.1 O léxico referente aos sufixos

4.2 O léxico referente aos prefixos

5. Palavras «não analisáveis»

5.1 Tipologia e estatística

5.2 Análise morfológica heurística

5.3 Probabilidades de categorias gramaticais no módulo heurístico

6. A razão de ser das expressões multipalavra

7. Esperanças para o futuro

Capítulo 6. As Morfolimpiadas e a avaliação da verificação ortográfica

Ricardo Ueda Karpishek

1. O projeto br.ispell e as Morfolimpiadas
2. O problema da verificação ortográfica
3. Complexidade do problema da verificação ortográfica
4. Limitações impostas pelo dicionário de verificação ortográfica
5. Interpretação das medidas das Morfolimpiadas
6. O que poderia ser um dicionário com um conteúdo mais rico
7. As formas desviantes

Capítulo 7. Unidades lexicais multpalavra, um osso duro de roer

Elisabete Marques Ranchhod e Cristina Mota

1. Introdução
2. Apresentação geral
3. A lista dourada
4. Participação na competição
 - 4.1 O pré-processamento
 - 4.2 A análise lexical
 - 4.3 Pós-processamento
5. Observações finais

Capítulo 8. Jspellando nas Morfolimpiadas

José João Almeida e Alberto Simões

1. Introdução ao Jspell
 - 1.1 Descrição dos dicionários Jspell
 - 1.2 Modos de funcionamento
 - 1.3 Não queremos as palavras todas!
 - 1.4 Programação usando Jspell
2. Participação na avaliação conjunta
 - 2.1 Avaliação: comentários gerais
 - 2.2 Derivação
 - 2.3 Como avaliar morfologia com derivação?
 - 2.4 Frequências
3. Conclusões

Capítulo 9. Radicalizadores versus analisadores morfológicos

Viviane Moreira Orengo e Diana Santos

1. Usos e história dos radicalizadores
2. Comparação com analisadores morfológicos
3. Avaliação de radicalizadores
4. O Removedor de Sufixos da Língua Portuguesa
5. Os resultados nas Morfolimpiadas
 - 5.1 Contabilização do resultado do sistema
 - 5.2 Comparação em termos dos grupos obtidos no corpus das Morfolimpiadas
 - 5.3 Comparação em termos do tamanho dos grupos
 - 5.4 Comparação usando o método de Paice
 - 5.5 Análise das semelhanças e diferenças entre os sistemas
6. Conclusões 87

Capítulo 10. Construção da lista dourada para as primeiras Morfolimpíadas do português

Anabela Barreiro e Susana Afonso

1. Introdução
2. Lista dourada: processo de construção
 - 2.1 Algumas questões problemáticas
 - 2.2 Directivas de revisão definidas pela organização
 - 2.3 Formato de codificação
3. Revisão de formas: problemas e soluções
 - 3.1 Multiplicidade associada às formas
 - 3.2 Lema
 - 3.3 Outros casos
4. Conclusão e sugestões para futuras morfolimpíadas para o português

Capítulo 11. Alguns comentários sobre a lista dourada

Jorge Baptista

1. Introdução
2. Lista dourada – breve apresentação
3. Decisões para simplificação das análises
4. Categorias gramaticais: ambiguidade, raridade
 - 4.1 Categorias gramaticais, definição de conjunto de etiquetas e granularidade
 - 4.2 Ambiguidade e raridade
5. Objectos textuais não identificados
6. Análise morfológica
 - 6.1 Palavras simples
 - 6.2 Palavras compostas
 - 6.3 Invenções
7. Conclusão

Capítulo 12. Avaliação conjunta de recuperação de informação da web portuguesa

Mário J. Silva, Bruno Martins e Miguel Costa

1. Introdução
2. Avaliação de sistemas de recuperação de informação
 - 2.1 «Webtrack»
3. Metodologia de avaliação
4. Pontuação
5. Considerações sobre a tarefa de avaliação
6. Justificação da abordagem proposta
7. Observações finais

Capítulo 13. CLEF: Abrindo a porta à participação internacional em avaliação de RI do português

Paulo Rocha e Diana Santos

1. Apresentação
2. O que é o CLEF?
 - 2.1 Participar na organização de uma avaliação conjunta internacional
 - 2.2 Comparando o CLEF e as Morfolimpíadas
3. O que significou adicionar o português
 - 3.1 A colecção
 - 3.2 Tópicos para recolha de informação (RI)

- 3.3 Perguntas para resposta automática a perguntas (RAP)
- 3.4 Avaliação de RI
- 3.5 Avaliação de RAP

4. Alguns comentários finais e balanço

- 4.1 Sugestões de melhoria para avaliação de RAP
- 4.2 Evolução e balanço da presença do português no CLEF

Capítulo 14. Avaliação de reconhecimento de entidades mencionadas: princípio de AREM

Cristina Mota, Diana Santos e Elisabete Ranchhod

- 1. Introdução
- 2. O que são entidades mencionadas e o que se entende pelo seu reconhecimento?
- 3. Por que razão é preciso identificar EM?
 - 3.1 Processamento sintáctico e semântico de texto
 - 3.2 Recolha de informação
 - 3.3 Resposta automática a perguntas
 - 3.4 Síntese de fala
 - 3.5 Geração de texto
 - 3.6 Tradução automática
- 4. Abordagens de reconhecimento de entidades mencionadas
- 5. Avaliação de reconhecimento de entidades mencionadas
- 6. Actividades preparatórias para avaliação conjunta de REM em português
- 7. Discussão no Avalon 2003 e desenvolvimentos futuros

Capítulo 15. Avaliação de tradução automática: alguns conceitos e reflexões

Luís Sarmiento, Anabela Barreiro, Belinda Maia e Diana Santos

- 1. Avaliação de tradução
- 2. Conceitos chave sobre avaliação de tradução automática
 - 2.1 Avaliação interna e avaliação externa
 - 2.2 Avaliação manual e automática
- 3. Algumas questões sobre a qualidade da TA
 - 3.1 Qualidade em função do objectivo
 - 3.2 Qualidade relativa à tradução humana
 - 3.3 Qualidade por correcção formal
 - 3.4 Qualidade em função do esforço de pós-edição
- 4. A via do pólo do Porto da Linguateca

Capítulo 16. Ferramentas para experimentação, recolha e avaliação de exemplos de tradução automática

Luís Sarmiento

- 1. Introdução
- 2. METRA: MEta – TRadutor Automático
 - 2.1 Descrição do METRA
 - 2.2 Observações suscitadas pelo METRA
- 3. BOOMERANG 161
 - 3.1 Descrição do BOOMERANG
 - 3.2 Observações sobre o BOOMERANG
- 4. TrAva
 - 4.1 Descrição e modo de funcionamento do TrAva
 - 4.2 O sistema de classificação do TrAva

5. Considerações finais

Capítulo 17. Uma experiência de recolha de exemplos classificados de tradução automática de inglês para português

Belinda Maia e Anabela Barreiro

1. Linguística e tradução
2. O papel do material de teste na avaliação de TA
3. Algumas áreas de dificuldade para a tradução automática
 - 3.1 Homografia e polissemia
 - 3.2 O sintagma nominal
 - 3.3 O sintagma verbal
 - 3.4 Grupos lexicais
 - 3.5 Outros
4. Conclusões

Capítulo 18. Avaliação de alinhadores

Alberto Simões e José João Almeida

1. Introdução
2. Alinhamento à frase
 - 2.1 Avaliação baseada em corpora pré-segmentados
 - 2.2 Avaliação baseada em corpora não segmentados
 - 2.3 Construção e obtenção dos casos de teste
3. Alinhamento à palavra
 - 3.1 Avaliação de alinhamento palavra a palavra em textos paralelos
 - 3.2 Avaliação de dicionários probabilísticos de tradução
4. Conclusão

Capítulo 19. Avaliação de sistemas de recuperação e categorização de textos: métodos e aplicações

Marco Gonzalez, Leonardo C. Langie e Vera L. S. de Lima

1. Introdução
2. Avaliação de sistemas de RI
 - 2.1 A noção de relevância
 - 2.2 Metodologia de avaliação em RI
 - 2.3 Medidas
 - 2.4 Formatos de apresentação dos resultados
3. Avaliação de sistemas de CT
4. Construção de coleções de referência
 - 4.1 Folha-Ricol
 - 4.2 Folha-Hierarq
5. Exemplos de aplicação das metodologias de avaliação
 - 5.1 Exemplos de avaliação de sistemas de RI
 - 5.2 Exemplo de avaliação de sistemas de CT
6. Considerações finais

Capítulo 20. As avaliações atuais de sistemas de busca na Web e a importância do usuário

Rachel Virginia Xavier Aires e Sandra Maria Aluísio

1. Introdução
2. Tipos de avaliação em RI
3. Criação de conjuntos de teste

- 4. A evolução dos sistemas de busca na Web e sua conseqüência para as abordagens de avaliação
- 4.1 Algumas tendências nos sistemas de busca na Web
- 4.2 A trilha High Accuracy Retrieval from Documents (HARD)
- 5. Considerações sobre a avaliação de sistemas da web para português

Capítulo 21. Avaliação de sistemas interactivos de recuperação de informação em bases de texto jurídicas

Paulo Quaresma e Irene Rodrigues

- 1. Introdução
- 2. Arquitectura
 - 2.1 Gestão de interacções
 - 2.2 Recuperação de informação
- 3. O nível de recuperação de informação
 - 3.1 Bases de texto
 - 3.2 Agente recuperador de informação
 - 3.3 Agente linguístico
 - 3.4 Agente de agrupamento
- 4. O nível de gestão de interacções
 - 4.1 O agente gestor de interacções
 - 4.2 Os agentes dos utilizadores
- 5. Avaliação do sistema de gestão das interacções
 - 5.1 Algoritmo de avaliação de listas de sugestões ao utilizador
 - 5.2 Convergência
 - 5.3 Testes realizados
- 6. Conclusões

Capítulo 22. A coleção TeMário e a avaliação de sumarização automática

Lucia Helena Machado Rino e Thiago Alexandre Salgueiro Pardo

- 1. Introdução
- 2. As principais abordagens para a construção de sistemas de SA
- 3. A avaliação de sistemas de SA
- 4. Tipos de avaliação
 - 4.1 Avaliações intrínsecas
- 5. Avaliações extrínsecas
 - 5.1 Categorização de documentos
 - 5.2 Recuperação de informação
 - 5.3 Perguntas e respostas
- 6. A construção do TeMário como uma coleção de avaliação de sistemas de SA
- 7. Discussão e conclusões

Capítulo 23. WPT 03: a primeira colecção pública proveniente de uma recolha da web portuguesa

Nuno Cardoso, Bruno Martins, Daniel Gomes e Mário J. Silva

- 1. Introdução
- 2. Criação da WPT 03
 - 2.1 Domínios abrangidos
 - 2.2 Tipos de documentos
 - 2.3 Sítios visitados
 - 2.4 Tamanho dos documentos
 - 2.5 Problemas encontrados

3. Outras colecções web e trabalhos relacionados
4. Caracterização da WPT 03
 - 4.1 Tamanho dos documentos
 - 4.2 Distribuição dos termos nos documentos
 - 4.3 Termos mais frequentes
5. O uso da WPT03 na avaliação e na investigação

Capítulo 24. Floresta Sintá(c)tica: ficção ou realidade?

Eckhard Bick, Diana Santos, Susana Afonso e Raquel Marchi

1. Apresentação
 2. Usos e razões de uma Floresta
 3. O recurso Floresta Sintá(c)tica
 - 3.1 Historial e resumo da metodologia
 - 3.2 Problemas de interpretação
 4. Avaliação conjunta usando a Floresta
 - 4.1 Avaliações usando florestas
 - 4.2 Passos iniciais de reconversão da Floresta
 5. Chamada às armas para participar na construção da Floresta
- Referências