

Segmentação topical automática de atas de reunião

Ovídio José Francisco
ovidiojf@gmail.com

RESUMO

Keywords

1. INTRODUÇÃO

Frequentemente atas de reunião tem a característica de apresentar um texto com poucas quebras de parágrafo e sem marcações de estrutura, como capítulos, seções ou quaisquer indicações sobre o tema do texto.

A tarefa de segmentação textual consiste dividir um texto em partes que contenham um significado relativamente independente. Em outras palavras, é identificar as posições onde há uma mudança significativa de tópicos.

É útil em aplicações que trabalham com textos sem quebras de assunto, ou seja, não apresentam parágrafos, seções ou capítulos, como transcrições automáticas de áudio e grandes documentos que contêm assuntos não idênticos como atas de reunião e notícias.

O interesse por segmentação textual tem crescido em em aplicações voltadas a recuperação de informação e sumarização de textos. Essa técnica pode ser usada para aprimorar o acesso a informação quando essa é solicitada por um usuário por meio de uma consulta, onde é possível oferecer porções menores de texto mais relevante ao invés de exibir um documento maior que pode conter informações menos pertinente. A sumarização de texto também pode ser aprimorada ao processar segmentos separados por tópicos ao invés de documentos inteiros.

Os algoritmos avaliados baseiam-se na ideia de coesão léxica entre assuntos. Isto é, a mudança de tópicos é acompanhada de uma proporcional mudança de vocabulário. A partir disso, vários algoritmos foram propostos. Nesse artigo, os principais serão analisados na perspectiva de atas de reunião.

2. TRABALHOS RELACIONADOS

2.1 Avaliação

Os entre os principais trabalhos relacionados a segmenta-

ção textual estão o *TextTiling* e o *C99*

avaliação todos precisam de um gold text

1 - Concatenação 2 - Juizes concordam ou não 3 - Mediador na reunião 4 - Não avaliar o segmentador e sim o resultado da aplicação final.

De acordo com [?] há duas principais dificuldades na avaliação de segmentadores automáticos.

3. ANÁLISE DOS RESULTADOS

4. TEXTTILINGBR

5. CONCLUSÃO

6. REFERENCES