

===== Review =====

*** Comentários para os Autores (Comentários para os Autores do Artigo. Deve conter a contribuição principal, pontos fortes e pontos a serem melhorados, e demais recomendações específicas. (preenchimento obrigatório)): O artigo apresenta um método de segmentação de documentos (atas de reuniões) em blocos por afinidade de conteúdo.

A aplicação é interessante tanto do ponto de vista teórico como prático, e o texto é bem escrito. Entretanto, o trabalho parece estar em uma fase inicial de desenvolvimento, atualmente concentrado na implementação de algoritmos já existentes, e sua aplicação ao processamento de textos em português (e não em inglês, como originalmente previsto). Não houve, até onde pude perceber, nenhuma proposta de melhoria ou adaptação destes recursos, mas simplesmente o reuso destes em um novo domínio e idioma.

O ponto que exige mais atenção é o trabalho de avaliação, que no estágio atual é extremamente reduzido, tanto pela quantidade de documentos considerados (seis atas) como pelo volume de informações em cada um (entre 10 e 32 sentenças cada uma). Torna-se assim difícil estimar se os resultados apresentados são de fato uma tendência geral, ou apenas um efeito observado neste pequeno conjunto de dados.

Falando sobre resultados, seria muito útil apresentá-los de uma forma mais concisa e que facilitasse a comparação entre os métodos. Em especial, as tabelas 3 e 4 poderiam ser combinadas em uma só, trocando linhas por colunas, de modo que a comparação entre TextTiling e C99 seja imediata.

===== Review =====

*** Comentários para os Autores (Comentários para os Autores do Artigo. Deve conter a contribuição principal, pontos fortes e pontos a serem melhorados, e demais recomendações específicas. (preenchimento obrigatório)): O artigo investiga técnicas de segmentação de atas de reunião por tópico abordado, focando-se na língua portuguesa. Apesar da relevância e ineditismo da tarefa, há alguns pontos problemáticos no artigo, principalmente quanto a sua avaliação. Apresento essas e outras preocupações abaixo, assim como algumas sugestões para melhoria.

Sugiro aos autores que, desde o início do artigo e também no título, adotem a terminologia mais explícita para o que fazem, que seria segmentação topical. Segmentação textual, usada de forma genérica, pode se referir a qualquer tipo de segmentação, e normalmente é entendida como segmentação sentencial.

A coleção de documentos utilizada é muito pequena. Sugiro aos autores que busquem outros documentos, de outros órgãos colegiados, para aumentarem sua base de dados. Nesse aspecto, uma questão importante que fica, e que não é comentada no artigo, é: a variedade de assuntos tratados nesse tipo de documento afeta os resultados? Em geral, nos trabalhos mais tradicionais, apesar da variedade de subtópicos, o tópico principal dos textos é mantido. No caso de atas, não há um único tópico principal, mas diversos. É importante incorporar esse tipo de discussão no artigo e, se possível, mensurar, de alguma forma, o impacto disso. A tarefa se torna mais simples ou mais difícil em função disso?

Obter mais documentos implica também em maior esforço na segmentação de referência. Uma estratégia simples que talvez pudesse ser utilizada para semi-automatizar a tarefa é realizar o alinhamento das atas com as convocações para as reuniões, em que os itens de pauta costumam ser listados de forma ordenada (e, normalmente, também são abordados nessa ordem nas atas).

Se entendi corretamente a descrição dos dados, a diferença entre o número de segmentos identificados em cada ata por cada participante é muito grande. Isso chegou a ser manualmente verificado? Provavelmente, os participantes adotaram pressupostos diferentes para a tarefa, o que acaba por não permitir uma avaliação robusta das técnicas investigadas.

Além disso, os autores afirmam que não é necessário que os limites entre os segmentos (real e hipótese) sejam idênticos, mas que se assemelhem em localização e quantidade. Não concordo com essa afirmação, pois, apesar da segmentação topical envolver de fato alguma subjetividade, deve se ter uma referência. É por essa razão que alguns dos trabalhos da literatura adotam os segmentos indicados por uma maioria dos anotadores, tendo-se, assim, uma listagem sobre quais são os segmentos com consenso e que deveriam minimamente ser detectados pelos algoritmos.

Nos resultados de avaliação, como se tratou a questão das segmentações divergentes entre os anotadores? Tirou-se a média entre elas, simplesmente? Se sim, isso parece inadequado, pois os anotadores claramente tiveram intenções diferentes em suas anotações de referência. Seria mais prudente apresentar as avaliações separadas para cada anotador. Do jeito que está, os resultados podem não permitir qualquer conclusão.

Sugiro aos autores que investiguem técnicas mais recentes de segmentação. Os dois trabalhos que abordam são importantes na área, mas há muita coisa nova surgindo e que pode ser de interesse dos autores.

Por fim, senti falta de exemplos no texto, tanto de trechos de atas (para o leitor ter ideia da natureza dos dados) quanto de exemplos segmentados corretamente e de forma errada. Isso permite ter uma ideia melhor das potencialidades e limitações dos métodos.

O artigo precisa de uma revisão cuidadosa da escrita. Em vários momentos, há problemas de estruturação sentencial.

==== Review =====

*** Comentários para os Autores (Comentários para os Autores do Artigo. Deve conter a contribuição principal, pontos fortes e pontos a serem melhorados, e demais recomendações específicas. (preenchimento obrigatório)): O trabalho é muito bom, e foi bem conduzido. Porém precisa de ajustes. Vou indicar algumas sugestões de melhorias aqui. Contudo, gostaria de comentar antes sobre o trabalho em si.

1) Segundo os autores, os algoritmos se baseiam, sobretudo, em frequência de termos. Porém, na seção 3, informam que a entonação e as pausas longas podem ser consideradas. Como isso é feito? a partir de marcas nas transcrições dos textos falados?

2) seria interessante levar em conta marcadores textuais (e.g., A propósito, Contudo, etc)

3) Problema com os experimentos: Como vocês lidaram com as diferenças na avaliação manual dos dois participantes? Olhando os valores indicados na Tabela 2 as discrepâncias são significativas. Contudo, isso não foi esclarecido na apresentação dos resultados dos experimentos. Vocês usaram a média aritmética entre os valores fornecidos por cada participantes? Isso é muito importante, pois levanta dúvidas sobre a qualidade dos resultados.

Ajustes no texto:

1) Faltam duas referências importantes: Friedman e Nemenyi.

2) Nomenclatura: os autores iniciam o documento falando sobre segmentação do texto em ASSUNTOS diferentes, e depois passam a usar o termo TÓPICO, que é o padrão na literatura.

3) A explicação sobre o alg C99 está confusa, apesar das figuras.

4) Revisar vírgulas, pois existem muitos usos incorretos dessa pontuação (por exemplo: Essa referência, deve / em que, pessoas capacitadas / novo, nesse (aqui seria ponto ou ;) / As medidas de avaliação tradicionais, podem não ser confiáveis).

5) Corrigir os números erros de concordância, principalmente referentes ao uso de voz passiva sintética (exemplos: escolheu-se os valores / atribuiu-se os valores / Calculou-se as medidas / computou-se também as medidas). existem ainda outros erros de concordância, como "Normalmente esse tipo de busca exige a inserção de termos exatos e apresentam ao usuário".

6) existem outros erros no texto relativos ao uso de vírgulas, acentuação e à concordância de gênero e número, mas seria muita coisa para eu escrever aqui. Apenas façam uma boa revisão, pois o trabalho é muito bom.