

Visualização e Exploração de Sumários Multidocumento

Jader Bruno Pereira Lima **Thiago A. S. Pardo**
Instituto de Ciências Matemáticas e de Computação, USP, SP

1. Introdução

Aplicações computacionais capazes de tratar adequadamente a incrível quantidade de informação disponível atualmente, principalmente na web, têm se tornado cada vez mais necessárias. A sumarização automática multidocumento é uma dessas aplicações, na qual, a partir de um conjunto de documentos que versam sobre um mesmo assunto, produz-se um único sumário, ou resumo (Mani, 2001). Tão importante quanto o processo de produção de um sumário é a forma de exposição deste e as possibilidades de se explorá-lo disponíveis para o usuário de um sistema de sumarização multidocumento.

Em uma visão mais generalista, os usuários podem ser divididos em dois grupos, dado as necessidades de interação destes com o sumário. (Mani e Maybury, 1999). O primeiro grupo de usuários necessita de uma visualização mais ampla, a qual quantifica os dados, ao invés de detalhá-los. Entretanto, um segundo grupo de usuários necessita de uma abordagem mais detalhista, a qual necessita de técnicas de visualização que possibilitem a ampliação da quantidade de informação a ser exibida ao usuário. A partir desses dois grupos de usuários, diversas funcionalidades podem ser extraídas para serem implementadas em um sistema de sumarização multidocumento.

2. Objetivos

Neste projeto de pesquisa, propõe-se a investigação de algumas destas funcionalidades e, com base nisso, a proposta de uma interface visual de interação com o usuário, seu desenvolvimento e avaliação.

3. Métodos e Procedimentos

Este projeto de pesquisa pode ser dividido em três fases: a fase de busca de trabalhos correlatos presentes na literatura; a fase de análise das funcionalidades de visualização que estarão presentes em nosso sistema de sumarização automática; e a fase de implementação desse sistema.

Um sistema de sumarização automática pode ter inúmeras funcionalidades tanto simples quanto avançadas. Algumas das funcionalidades mais importantes analisadas neste trabalho foram: *link* em cada sentença para seu texto fonte; destaque da sentença escolhida em seu texto fonte; exibição das palavras chave do sumário; data de publicação dos textos fonte; parametrização do sistema de sumarização (como tamanho do sumário); e agrupamento das sentenças por tópicos com um mesmo significado.

4. Conclusões Parciais

Nesta fase do projeto, mais funcionalidades estão sendo levantadas, e, futuramente, pretende-se a implementação do sistema e o teste de sua funcionalidade em sumarizadores existentes (Ribaldo et al., 2012).

Referências Bibliográficas

- [1] Mani, I. (2001). Automatic Summarization. John Benjamins Publishing Co. Amsterdam.
- [2] Mani, I. and Maybury, M. T. (1999). Advances in automatic text summarization. MIT Press, Cambridge, MA.
- [3] Ribaldo, R.; Akabane, A.T.; Pardo, T.A.S. (2012). Multi-document Summarization with Graph Metrics. In the (on-line) *Proceedings of the PROPOR 2012 Demonstrations*, pp. 1-3.