

Uma interface para visualização e exploração de sumários multidocumento

Jader Bruno Pereira Lima, Thiago Alexandre Salgueiro Pardo

Instituto de Ciências Matemáticas e de Computação – ICMC

Universidade de São Paulo – USP

jbplima@usp.br, taspardo@icmc.usp.br

Objetivos

A sumarização automática multidocumento é definida como a tarefa de se produzir automaticamente um sumário a partir de vários textos fonte que versam sobre um mesmo assunto (Radev e Mckeown, 1998). Tão importante quanto o sumário produzido é a visualização das informações relacionadas, de forma que se permita ao usuário de um sistema de sumarização a exploração do conteúdo de interesse. Neste trabalho, visamos o estudo de sistemas de sumarização e, a partir disso, a implementação de uma interface flexível e amigável (chamada ViSUM) para sumarização.

Método

Para a realização deste trabalho, foram realizados estudos e análises sobre funcionalidades existentes em sistemas correlatos presentes na literatura, como o sistema Columbia Newsblaster (Mckeown et al., 2002), que é um dos mais utilizados e difundidos da área. As funcionalidades levantadas foram analisadas com relação a algumas características, por exemplo, o tipo de usuário que é o seu foco, separando-os em usuários casuais e avançados. Além disso, em função das funcionalidades presentes na literatura, identificaram-se desde funcionalidades essenciais a opcionais.

Na Figura 1 temos a interface do sistema ViSUM – Sistema de Visualização de Sumários Multidocumento, na qual está sendo apresentado um sumário juntamente com os seus textos fonte originais. Alguns exemplos de funcionalidades implementadas são a referência da sentença selecionada no sumário ao seu texto fonte, por meio de coloração da mesma, e apresentação de imagens ilustrativas sobre o assunto do sumário.

Este sistema tem o objetivo de ser genérico, ou seja, poderá ser utilizado por qualquer sistema

de sumarização multidocumento, de acordo com a preferência do usuário.

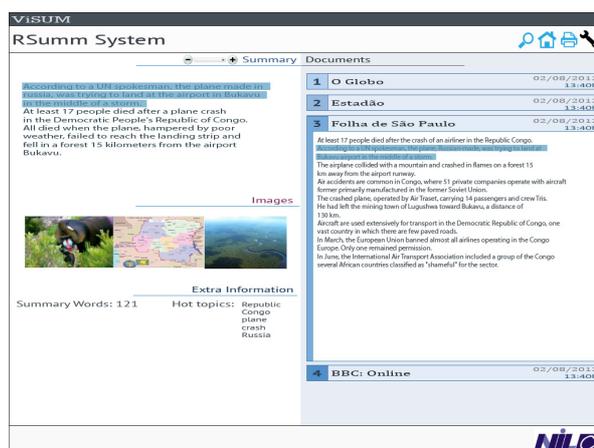


Figura 1: Interface do Sistema ViSUM

Conclusões e Atividades Futuras

Este trabalho se encontra na fase de implementação da interface. As atividades seguintes serão a finalização da implementação e a sua avaliação, que será feita com usuários reais e métricas clássicas da área de Interação Humano-Computador.

Agradecimentos

À FAPESP, pelo apoio a este trabalho.

Referências Bibliográficas

- McKeown, K. R.; Barzilay, R.; Evans, D.K.; Hatzivassiloglou, V.; Klavans, J.L.; Nenkova, A.; Sable, C.; Schiffman, B.; Sigelman, S. (2002). Tracking and summarizing news on a daily basis with columbia's newsblaster. In the *Proceedings of the Human Language Technology Conference*, pp. 280-285.
- Radev, D. R. and McKeown, K. (1998). Generating natural language summaries from multiple on-line sources. *Computational Linguistics*, vol. 24, N. 3, pp. 469-500.