UMA INTERFACE PARA VISUALIZAÇÃO E EXPLORAÇÃO DE SUMARIOS MULTIDOCUMENTO

Jader Bruno Pereira Lima, Thiago Alexandre Salgueiro Pardo
Instituto de Ciências Matemáticas e de Computação (ICMC), USP/São Carlos
Núcleo Interinstitucional de Linguística Computacional
jbplima@usp.br, taspardo@icmc.usp.br

RESUMO

Neste artigo, são investigadas funcionalidades de visualização e exploração de sumários multidocumento presentes em vários sistemas da literatura, e, a partir deste estudo, é proposto um sistema de visualização de sumários multidocumento, com o objetivo de melhorar a experiência de leitura dos usuários deste tipo de sistema.

1. INTRODUÇÃO

Com o crescente aumento da quantidade de informação disponibilizada atualmente, principalmente na web, faz-se necessário o desenvolvimento e pesquisa de técnicas computacionais que sejam capazes de tratar essas informações. Neste contexto, está inserida a Sumarização Automática (SA), que é uma subárea da Inteligência Artificial. A SA consiste na produção de uma versão mais curta de um texto fonte contendo as suas informações mais relevantes (Mani, 2001), o qual é chamado de sumário, ou resumo. Este trabalho está inserido em uma vertente de pesquisa relacionada à SA, que visa a criação de sumários a partir de vários textos que versam sobre o mesmo assunto: a Sumarização Automática Multidocumento (SAM).

O resultado esperado de ambos os processos descritos acima, a SA e a SAM, é o sumário textual, para que o usuário possa adquirir o conhecimento de maneira rápida e resumida. Porém, quando focamos na SAM, é possível que sejam disponibilizadas mais informações sobre os dados sumarizados, além de apenas o sumário em si. Como o sumário é criado a partir de fragmentos extraídos de vários textos fonte, seria interessante se fosse possibilitado ao usuário o rastreamento desses textos fonte a partir de cada fragmento destes contidos no sumário. Isso enriqueceria ainda mais a experiência do usuário nessa tarefa de leitura do sumário. Além dessa funcionalidade de rastreamento, outras funcionalidades podem ser implementadas em um sistema de SAM, por exemplo, a exibição de imagens ilustrativas sobre o conteúdo do sumário, a possibilidade de parametrização do tamanho do sumário gerado, e a exibição das palavras chave do sumário.

Neste contexto, apresenta-se uma proposta de interface de visualização de sumários multidocumento.

2. MATERIAL E/OU MÉTODOS

Foram estudados vários sistemas de SAM com ênfase em suas funcionalidades de visualização. Para este estudo, foram levantados os principais sistemas existentes na área, os quais visam prover aos usuários uma experiência de leitura de sumários multidocumento mais rica. Estes sistemas são:

Columbia Newsblaster (McKeown et al., 2002), News in Essence (Radev et al., 2004), Ineats (Leuski et al., 2000), RSumm (Ribaldo et al., 2012) e SumView (Wang et al., 2012).

A Tabela 1, lista todas as funcionalidades encontradas nestes sistemas Nesta tabela, indica-se, para cada sistema, se a funcionalidade estava presente ou não, por exemplo, nota-se que a funcionalidade 1.1 "Referência aos textos-fonte" faz parte de todos os sistemas, exceto do sistema SumView.

3. RESULTADOS E DISCUSSÃO

A Figura 1 apresenta o protótipo da interface do sistema proposto (chamado de ViSUM – Visualizador de Sumários Multidocumento), o qual está em fase final de desenvolvimento neste momento da pesquisa. Nesta interface, o sumário está à esquerda e os textos fonte à direita, os quais são referenciados por meio da coloração de suas sentenças correspondentes ao se colocar o ponteiro do mouse sobre alguma sentença do sumário.

Outras funcionalidades são a exibição de imagens ilustrativas do sumário (recuperadas de uma API de buscas na web), a possibilidade de parametrização do tamanho do sumário, a exibição de informações extras sobre o sumário, por exemplo, as palavras mais frequentes e a exibição de palavras chave do sumário.

A escolha da inclusão dessas funcionalidades se seu pelo estudo realizado anteriormente, relatado na Seção 2, de sistemas com esta mesma finalidade. No sistema ViSUM, foram incluídas todas as funcionalidades consideradas indispensáveis ou essenciais, ou seja, todas que estão presentes em pelo menos 3 dos 5 sistemas estudados.

4. CONCLUSÕES PARCIAIS

Nesta etapa da pesquisa, o sistema está sendo finalizado e, como próximo passo, será feita uma avaliação com usuários reais, seguindo métricas de usabilidade presentes na literatura.

AGRADECIMENTOS

À FAPESP, pelo apoio a este projeto.

REFERÊNCIAS

Leuski, A; Lin, C.Y.; Hovy, E.H. (2003). iNeATS: Interactive Multidocument Summarization. In the *Proceedings of 41st Annual Meeting on Association for Computational Linguistics*, pp. 125-128.

Mani, I. (2001). *Automatic Summarization*. John Benjamins Publishing Co., Amsterdam.

McKeown, K.R.; Barzilay, R.; Evans, D.K.; Hatzivassiloglou, V.; Klavans, J.L.; Nekova, 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 second international conference on Human Language Technology Research*, pp. 280-285.

Radev, D.; Jing, H.; Syvs, M.; Tam, D. (2004). Centroid-based summarization of multiple documents. *Information Processing and Management*, Vol. 40, pp. 919-938.

Ribaldo, R.; Akabane, A.T.; Rino, L.H.M.; Pardo, T.A.S. (2012). Graph-based Methods for Multi-document Summarization: Exploring Relationship Maps, Complex Networks and Discourse Information. In the *Proceedings of the 10th International Conference on Computational Processing of Portuguese* (LNAI 7243), pp. 260-271.

Wang, D.; Li, T.; Zhu, S. (2013). SumView: A web-based engine for summarizing product reviews and costumer opinions. *Expert Systems and Applications*, Vol. 40, N. 1, pp. 27-33.

Tabela 1: Funcionalidades de visualização encontradas nos sistemas de SAM estudados

	Columbia Newsblaster	News in Essence	Ineats	RSumm	SumView
Grupo 1: Visualização de Informação					
1.1: Referência aos textos fonte	Sim	Sim	Sim	Sim	Não
1.2: Destaque da sentença escolhida em seu texto-fonte	Sim	Sim	Não	Não	Não
1.3: Adequação das informações textuais (remoção de *tags* HTML, por exemplo)	Sim	Não	Sim	Sim	Sim
1.4: Exibição das palavras chave do sumário	Sim	Não	Sim	Não	Sim
1.5: Exibição da quantidade de palavras dos textos-fonte	Sim	Não	Sim	Não	Sim
1.6: Exibição da data de publicação dos textos-fonte	Sim	Não	Não	Não	Não
1.7: Exibição do local de publicação dos textos-fonte	Não	Não	Sim	Não	Não
Grupo 2: Funcionalidades avançadas de visualização					
2.1: Assuntos mais sumarizados pelos usuários	Não	Sim	Não	Não	Não
2.2: Exibição de todos os textos-fonte utilizados	Não	Sim	Não	Sim	Não
2.3: Exibição de imagens ilustrativas do sumário	Sim	Não	Não	Não	Sim
2.4: Diferenciação das sentenças por cores de acordo com seu tópico ou texto-fonte	Não	Não	Sim	Não	Não
2.5: Agrupamento dos textos-fonte por tópicos	Sim	Não	Não	Não	Sim
Grupo 3: Parametrização do sumarizador					
3.1: Parametrização do tamanho do sumário	Não	Sim	Sim	Sim	Não
3.2: Escolha dos tópicos mais relevantes a serem sumarizados	Não	Não	Sim	Não	Sim

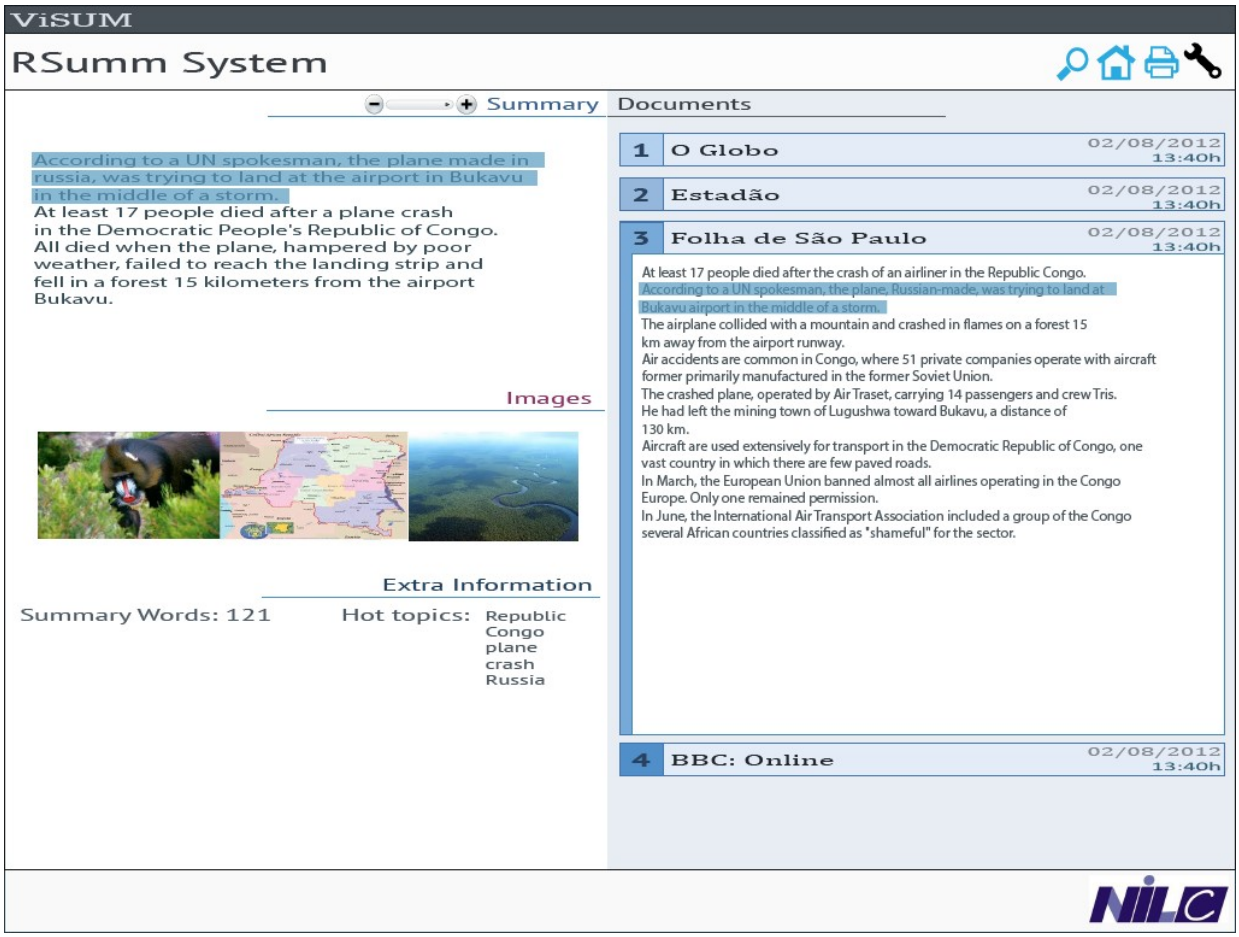

Figura 1: Interface do sistema ViSUM