

# Sumarização Multidocumento com Base em Métricas de Redes Complexas

Ademar T. Akabane<sup>1</sup>, Thiago A. S. Pardo<sup>1</sup>, Lucia H. M. Rino<sup>2</sup>

<sup>1</sup>Instituto de Ciências Matemáticas e de Computação, USP, São Carlos-SP

<sup>2</sup>Departamento de Computação, UFSCar, São Carlos-SP

## Objetivos

Com a crescente popularização do acesso à internet, percebe-se que houve um aumento exponencial das informações geradas e distribuídas em meios digitais. Neste contexto, a Sumarização Automática (SA) multidocumento, que consiste na produção de um resumo a partir de vários textos/documentos que versam sobre um mesmo assunto (Mani, 2001), está obtendo maior relevância para tentar amenizar o problema decorrente do excesso de informação disponível.

O objetivo deste trabalho foi estudar e aplicar algumas métricas de Redes Complexas (RC) na SA multidocumento, para extrair as informações mais relevantes dos textos-fonte. As métricas adotadas foram: Grau (Gr), Caminho Mínimo (C.M.), Coeficiente de Aglomeração (C.A.) e uma combinação (Comb) dessas medidas. Essas medidas indicam o grau de conectividade entre nós da rede.

## Métodos e Procedimentos

Neste projeto, os textos foram representados em uma RC. Cada nó da rede é uma unidade sentencial e as arestas as relações/pesos entre pares de sentenças distintas. Esses pesos foram calculados a partir da medida do cosseno, de Salton et al. (1997), ou seja, quanto maior o peso, maior será a similaridade entre as sentenças. Nesse cálculo, as *stopwords* não são consideradas por não adicionarem informatividade.

As métricas empregadas nesse trabalho associam um valor a cada nó da rede, dando, assim, a condição de escolha das sentenças que possivelmente irão compor o sumário.

No desenvolvimento desse projeto, teve-se apoio de outros materiais/ferramentas, como: a) um corpus de textos e resumos; b) um segmentador textual; c) um radicalizador; d) ferramenta ROUGE (Lin e Hovy, 2003), para

avaliação de sumários (comparando-os com sumários humanos).

## Resultados

Os resultados obtidos a partir da avaliação da ferramenta ROUGE são dados na Tabela 1 em termos das medidas clássicas de precisão, cobertura e medida-f. Os valores da tabela variam entre 0 e 1, onde 0 indica que o sumário automático é diferente do sumário humano e 1 indica a proximidade máxima entre ambos.

Tabela 1: Tabela dos resultados das avaliações de cada métrica proposta

	Precisão	Cobertura	Medida-f
<b>Gr</b>	0.533	0.504	<b>0.515</b>
<b>C.A.</b>	0.467	0.447	0.456
<b>C.M.</b>	0.530	0.500	0.513
<b>Comb</b>	0.528	0.506	0.514

Pela análise da Tabela 1, observa-se que a medida que resultou no melhor desempenho, ou seja, aquela que gerou melhores sumários, corresponde à métrica mais simples, dentre todas as métricas propostas neste projeto, a de Grau.

## Agradecimento

Ao PIBIC/CNPq, pela concessão da bolsa de Iniciação Científica.

## Referências Bibliográficas

- Lin, C.Y. and Hovy, E. (2003). Automatic Evaluation of Summaries Using N-gram Cooccurrence Statistics. In *Proceedings of 2003 Language Technology Conference (HLTNAACL 2003)*, Edmonton, Canada.
- Mani, I. (2001). *Automatic Summarization*. John Benjamins Publishing Co., Amsterdam.
- Salton, G.; Singhal A.; Mitra, M; Buckley C. (1997). Automatic Text Structuring And Summarization. *Information Processing & Management*, Vol. 33, No, 2, pp. 193-207.