

Eventos, 1º Colóquio de Linguística para o Processamento Automático de Linguagem Natural

[CAPA](#) [SOBRE](#) [ACESSO](#) [CADASTRO](#) [PESQUISA](#) [EDIÇÕES ANTERIORES](#)

Capa > Colóquio de Linguística para o Processamento Automático de Linguagem Natural > 1º Colóquio de Linguística para o Processamento Automático de Linguagem Natural > Submissões gerais > **Tironi de Camargo**

Tamanho da fonte:

Anotação e Descrição de Corpus para a Investigação de Estratégias de Sumarização Humana Multidocumento
Renata Tironi de Camargo, Ariani Di Felippo

Prédio: Prédio ED III
Sala: Sala 302
Data: 2012-11-29 06:20 – 06:50
Última alteração: 2012-11-01

Resumo

A Sumarização Automática Multidocumento (SAM) é uma das várias aplicações desenvolvidas no Processamento Automático das Línguas Naturais (PLN). Nela, busca-se, em última instância, desenvolver sistemas capazes de produzir um sumário (resumo) a partir de uma coleção de textos-fonte que tratam de um mesmo assunto. Tendo em vista que o interesse pela SAM é relativamente recente em comparação à sumarização automática monodocumento, os pesquisadores do PLN não dispõem de estudos sistemáticos a respeito da sumarização humana multidocumento (SHM) que possam subsidiar a SAM. Diante desse cenário, estabeleceu-se o objetivo geral de se caracterizar sumários humanos multidocumento em PB com vistas à identificação de estratégias de seleção de conteúdo que possam subsidiar a SAM. Para tanto, delimitaram-se as seguintes tarefas: (i) seleção e anotação de *corpus*, (ii) caracterização de sumários multidocumento, (iii) identificação de estratégias de seleção de conteúdo, (iv) formalização das estratégias de seleção de conteúdo e (v) avaliação das estratégias formalizadas. Neste trabalho, descrevem-se, especificamente, as tarefas de (i) seleção e anotação de *corpus* e (ii) caracterização de sumários multidocumento.

Palavras-chave

Sumarização