



SALÃO DE INICIAÇÃO CIENTÍFICA XXVIII SIC

paz no plural



Evento	Salão UFRGS 2016: SIC - XXVIII SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2016
Local	Campus do Vale - UFRGS
Título	RecETC: um foco no aprimoramento da Mineração Textual do Recomendador do ETC
Autor	NATANAEL RODRIGO ZIMMER
Orientador	PATRICIA ALEJANDRA BEHAR

RecETC: um foco no aprimoramento da Mineração Textual do Recomendador do ETC

Aluno: Natanael Rodrigo Zimmer

Orientador: Patricia Alejandra Behar

UFRGS – Universidade Federal do Rio Grande do Sul

O RecETC (Recomendador do ETC) refere-se a um Sistema de Recomendação integrado ao ETC (Editor de Texto Coletivo). O ETC foi desenvolvido pelo NUTED (Núcleo de Tecnologia Digital Aplicada à Educação) visando oferecer um ambiente para a Escrita Coletiva. Com o propósito de favorecer apoio a esse processo é que o RecETC foi idealizado. O mesmo possui como objetivo recomendar materiais, em diferentes formatos, com base no desenvolvimento da produção textual coletiva. Para receber recomendações, os usuários devem dar início à construção do texto e clicar sobre o botão “RecETC”. No entanto, a partir de um processo de validação, identificou-se a necessidade de aprimoramento do minerador de texto, uma vez que as palavras-chave mineradas não estavam representando com propriedade relações com o assunto dos textos que estavam sendo produzidos no ETC. Logo, foram realizados estudos acerca das temáticas de Mineração de Textos e Sistemas de Recomendação visando identificar métodos para qualificar a extração do possível tema do texto, bem como em relação aos materiais recomendados.

Nesse sentido, as principais melhorias realizadas no Minerador de Textos referem-se a alteração do método de obtenção das palavras-chave para que o sistema consiga identificar o possível assunto que está sendo tratado no texto. Dessa forma, foi implementado um algoritmo capaz de diferenciar termos comuns de termos compostos por múltiplas palavras. Além da utilização da contagem da frequência de termos foi efetuada uma normalização das palavras-chave encontradas. Essa refere-se a transformação de palavras no plural para o singular, bem como sua reversão para o verbo no infinitivo com o intuito de identificar a efetiva quantidade de cada termo. Diante disso, fez-se uso de um método denominado Term Frequency - Inverse Document Frequency (TF-IDF) com o objetivo de registrar uma pontuação para a identificação do índice de relevância dos termos. Esses são implementados pelo Sistema de Gerenciamento de Banco de Dados (SGBD) MySQL utilizando técnicas de busca Full-Text Search. Com o objetivo de qualificar ainda mais o processo de recomendação, foi integrado à interface do usuário uma área para que o mesmo tenha autonomia para excluir ou incluir novas palavras-chave. Em caso de exclusão, o termo é destacado na cor vermelha e direcionado para uma área separada das palavras mineradas. As palavras incluídas manualmente pelo usuário são destacadas em azul também em área diferenciada. Nessa caso, elas são mineradas junto com as identificadas automaticamente pelo minerador. As palavras foram separadas em áreas visando destacar para o usuário as mudanças realizadas, bem como flexibilizar possíveis reversões identificadas como necessárias pelo usuário. É importante destacar que essas mudanças ocorrem apenas no texto no qual as recomendações estão sendo solicitadas, não influenciando nos demais textos e usuários do ETC. Além disso, identificou-se como relevante dar flexibilidade para que o usuário possa interagir com o RecETC visando refinar e especificar suas necessidades e interesses, caracterizando assim seu perfil no texto por meio das palavras-chave.

Com base nas melhorias realizadas, o RecETC está sendo utilizado com alunos do curso de Pós-Graduação em Informática na Educação da UFRGS a fim de validar os aprimoramentos sucedidos. Pretende-se com essa aplicação avaliar a qualidade da mineração do texto e as interações realizadas pelos usuários ao incluir/excluir novos termos, assim como as recomendações oriundas do processo de mineração. A partir dos resultados coletados poderão ser feitos novos ajustes e aperfeiçoamentos no RecETC.