

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



Aluno: Natanael Rodrigo Zimmer
Orientador: Patricia Alejandra Behar



O RecETC (Recomendador do ETC) refere-se a um Sistema de Recomendação integrado ao ETC (Editor de Texto Coletivo) que visa oferecer apoio ao processo de Escrita Coletiva. Tem como objetivo recomendar materiais, em diferentes formatos, com base no desenvolvimento da produção textual coletiva.

Desafio:

As palavras-chave mineradas não estavam representando a contento o assunto dos textos que estavam sendo produzidos no ETC

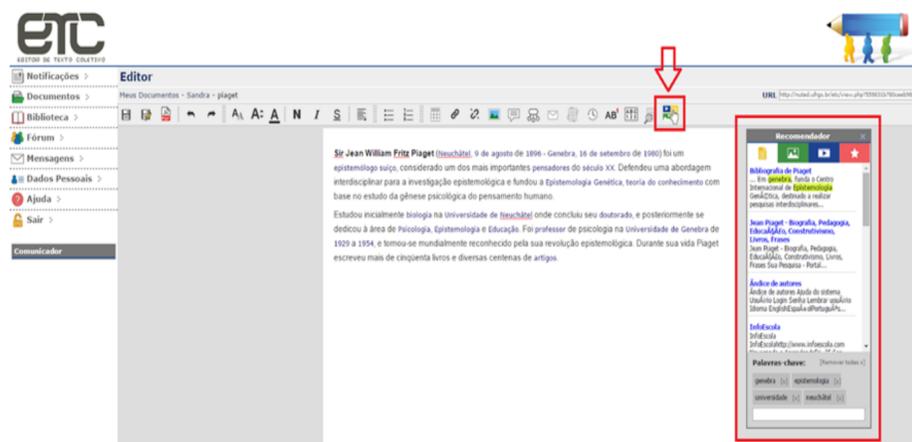
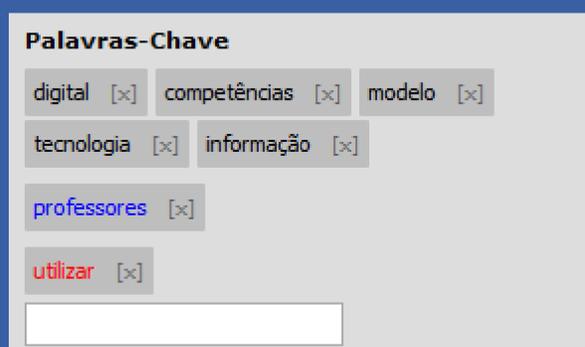
Solução:

Implementação de um algoritmo capaz de reconhecer termos compostos por múltiplas palavras e reconhecer diferentes variações de um mesmo termo como sendo iguais.

Outras melhorias no Minerador de Textos:

* Construção de área na interface do usuário para que o mesmo tenha autonomia para excluir ou incluir novas palavras-chave.

* Palavras excluídas são marcadas em vermelho e as adicionadas são destacadas em azul.



Abordagem:

Ao efetuar a contagem da frequência de termos foi feita uma normalização das palavras-chave encontradas, transformando 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.

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.

Para isso, foi utilizada a técnica de Full-Text Search que é implementada pelo SGBD MySQL.