

# Uso de Webservices no Apoio à Atividade de Mineração de Texto do Sobek

José Luis Damaren Junior <sup>1</sup>, Eliseo Berni Reategui <sup>2</sup>

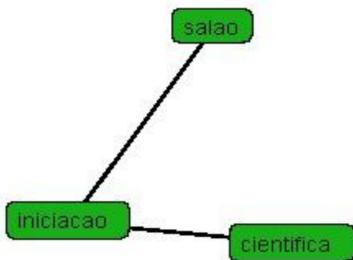
<sup>1</sup> Autor, Ciência da Computação, Universidade Federal do Rio Grande do Sul  
<sup>2</sup> Orientador



CH - Ciências Humanas

## SOBEK

O Sobek é uma ferramenta de extração de informação capaz de identificar conceitos relevantes em um texto. O método que usa é conhecido como mineração de texto. Através de uma análise estatística das palavras, é possível conseguir muita informação rapidamente.



A imagem acima é um exemplo de grafo gerado pelo aplicativo online do Sobek ([sobek.ufrgs.br](http://sobek.ufrgs.br)). Apesar do resultado ser satisfatório, queremos testar uma comunicação cliente-servidor com sockets TCP para o Sobek, que é diferente do que é realizado no momento.

## WEB CHAT

Um chat textual foi desenvolvido utilizando os sockets TCP para comunicação entre cliente e servidor. Todo o texto do chat será analisado pelo Sobek, e dessa maneira achamos uma possibilidade de se usar esse tipo de comunicação juntamente com o aplicativo. Abaixo está uma imagem mostrando a interface do chat.



## IMPLEMENTAÇÃO

Os sockets TCP foram utilizados pelo fato do protocolo ser confiável, o que significa que os pacotes, que carregam o conteúdo da comunicação, chegam sem duplicação, em ordem e sem erros. Caso sockets UDP ou alguma outra maneira de comunicação não confiável fosse utilizada, a aplicação deveria conferir a integridade de toda a informação.

O lado do cliente foi desenvolvido em Javascript, e o do servidor em C++. O servidor também guarda informações em um banco de dados MySQL. O chat apresenta, além da funcionalidade padrão de conversa, a possibilidade de pedir ajuda a um agente mecânico, que fica na direita da interface. Lá estão disponíveis botões para funcionalidades diversas, que agem sobre o texto digitado na caixa de texto exclusiva, também a direita.

## FONTES

Tanenbaum, Andrew S. (2003-03-17). Computer Networks (Fourth ed.). Prentice Hall. ISBN 0-13-066102-3.

[sobek.ufrgs.br](http://sobek.ufrgs.br)

[sobek.ufrgs.br/chat/chat.html](http://sobek.ufrgs.br/chat/chat.html)

[en.wikipedia.org/wiki/Transmission\\_Control\\_Protocol](http://en.wikipedia.org/wiki/Transmission_Control_Protocol)



MODALIDADE  
DE BOLSA

Iniciação Científica