



Evento	Salão UFRGS 2013: SIC - XXV SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2013
Local	Porto Alegre - RS
Título	Uso de Webservices no Apoio à Atividade de Mineração de Texto do Sobek
Autor	JOSÉ LUIS DAMAREN JUNIOR
Orientador	ELISEO BERNI REATEGUI

O Sobek é uma ferramenta de extração de conceitos capaz de identificar conceitos relevantes em um texto e apresentá-los ao usuário. O método de extração de conceitos utilizado pelo Sobek é conhecido como *mineração de texto*, que se difere da *mineração de dados* por trabalhar com dados não estruturados. O principal objetivo da ferramenta é facilitar a compreensão de um texto ou conjunto de textos. Para isso, realiza uma análise estatística sobre o texto, extraíndo os conceitos mais recorrentes e gerando uma base de conhecimento que pode ser treinada e aprimorada. Os principais conceitos do texto são expostos em um grafo, que demonstra sua relação e relevância. O objetivo do bolsista é aperfeiçoar a comunicação cliente servidor do site do aplicativo. No site, é possível utilizar o Sobek de forma online, inserindo-se um texto e obtendo-se o grafo referente a esse texto como resposta. No entanto, a forma como a comunicação é realizada no momento não é apropriada a todos os casos. A alternativa escolhida para aperfeiçoar a comunicação foi o uso de sockets TCP. Esse protocolo é conveniente pelo fato de ser confiável, ou seja, todos os pacotes que foram enviados chegam, sem erro e na ordem em que foram enviados. Nesse caso, o cliente e o servidor devem criar um socket e estabelecer uma conexão para poderem transmitir dados. Isso exige que o cliente já saiba a porta em que o servidor está operando e que o servidor esteja sempre pronto a receber novas requisições de conexão. A criação dessa comunicação permite que o Sobek seja utilizado junto com outros processos, de forma “invisível” ao usuário, e que seja mais amplamente divulgado no meio acadêmico.