

# Mineração de Texto Educacional

Mineração com um simples clique

SOBEK Versão 1.1



## Breve histórico

- 2007 Sobek foi criado como uma ferramenta de mineração de texto para auxiliar os professores do ensino a distância a avaliarem o trabalho dos alunos feito colaborativamente;
- 2009 Sobek começou a ser utilizado para auxiliar os alunos na compreensão da leitura e tarefas de resumo de texto;
- 2010 em diante Sobek foi incorporado em outros sistemas, tais como: avaliação de posts dos alunos em fóruns de discussão, jogos digitais para promover a narrativa escrita, ferramenta de aprendizagem baseada em projetos com recomendação de conteúdo.

## O que é mineração de texto?

Mineração de texto é uma técnica da Ciência da Computação que permite a análise e identificação de informações relevantes em textos.

Diferentes métodos podem ser usados para isso, desde métodos estatísticos até o processamento de linguagem natural. A ferramenta Sobek foi construída a partir de um algoritmo inicialmente definido por Schenker (2003) e subsequentemente modificado pelo Grupo de Pesquisa GTech.Edu a fim de torná-lo mais adequado às aplicações educacionais (Reategui et al., 2011).

Sobek é extremamente fácil de usar. Para ver uma representação visual de um texto você simplesmente tem que copiá-lo/colá-lo no editor da ferramenta e pressionar um botão. Com um simples clique você pode visualizar um diagrama conciso contendo os principais termos e relações do seu texto.



Prof. Eliseo Reategui (coordenador do GTech.Edu, no meio) e estudantes pesquisadores. Da Direita para a esquerda: Daniel Epstein, Isis da Costa Pinho, Otavio Acosta, Heli Meurer e Fabio Damasceno.

### REFERÊNCIAS

- Reategui, E., Epstein, D., Lorenzatti, A., Klemann, M. Sobek: a Text Mining Tool for Educational Applications. In: International Conference on Data Mining, 2011, Las Vegas, USA, 2011, 59-64.
- Schenker, A. Graph-Theoretic Techniques for Web Content Mining. PhD thesis, University of South Florida, 2003.

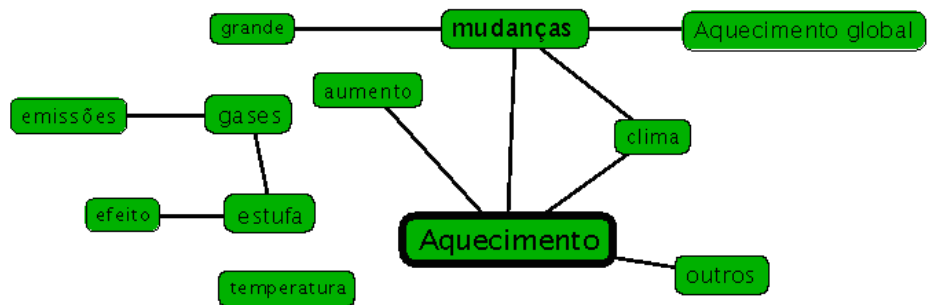


## Como funciona?

O Sobek procura por termos frequentes e suas relações em um texto

Ao analisar as palavras recorrentes em um texto, Sobek é capaz de identificar termos relevantes e representá-los graficamente, como mostra o exemplo à direita. O texto analisado é sobre o efeito do aquecimento global, obtido da Wikipédia.

No gráfico, os grandes nodos representam os termos mais frequentes e as conexões representam as relações entre eles.



Texto analisado: [http://pt.wikipedia.org/wiki/Aquecimento\\_global](http://pt.wikipedia.org/wiki/Aquecimento_global)

## Outras características

### UM, O, QUE, COM, ...

Palavras frequentes tais como artigos, preposições e advérbios são simplesmente descartadas.

### THESAURUS

Um thesaurus é utilizado para identificar conceitos que são referidos usando palavras diferentes, de modo que eles possam ser representados em um único nodo.



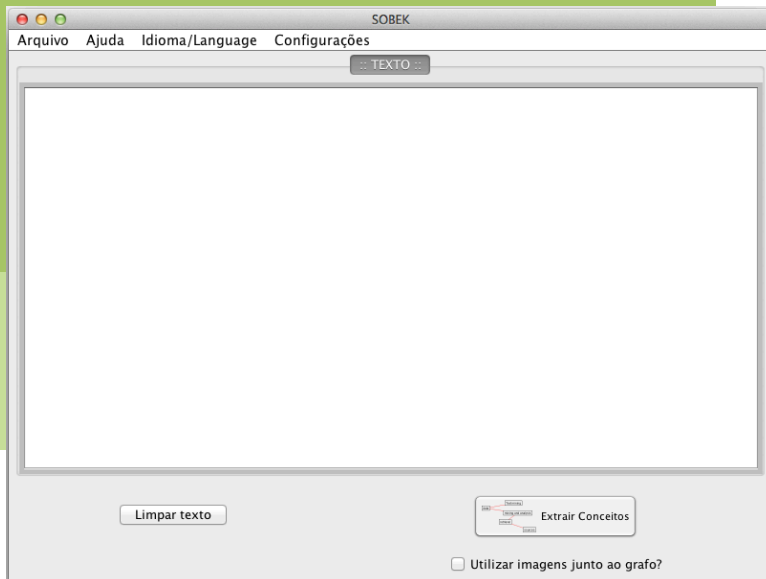
# Passo a Passo

## 1. Comece fazendo o download da ferramenta em: [sobek.ufrgs.br](http://sobek.ufrgs.br)

Você pode abrir o aplicativo .jar apenas clicando no arquivo baixado.

Você também pode executar o aplicativo online apenas acessando o site do Sobek e escolhendo esta opção.

Ao abrir o aplicativo, o sistema abrirá uma janela como a da esquerda. Ela funciona como um editor de texto no qual você pode digitar ou colar o texto que você deseja analisar.



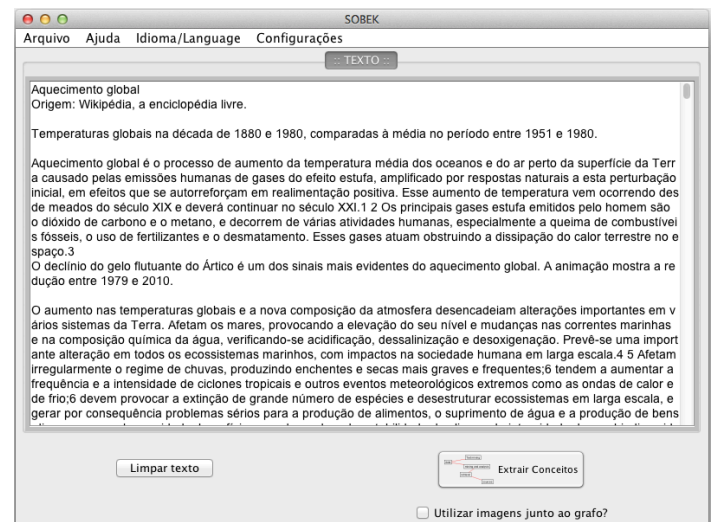
## 2. Copiar/colar um texto no editor do Sobek

Basta clicar no espaço em branco na tela principal do Sobek e colar o texto que você deseja analisar.

Você também pode digitar o texto ou, alternativamente, importá-lo usando o menu Arquivo no topo da tela

## 3. Pressione o botão **Extrair Conceitos**

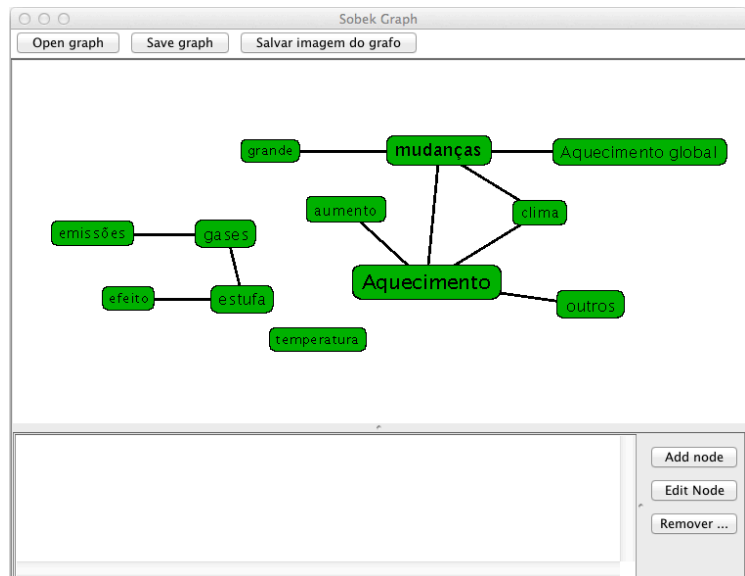
Ao clicar no botão Extrair Conceitos, no lado inferior direito da janela, uma pequena janela será exibida com a mensagem "Processando texto. Por favor, aguarde...". Geralmente, para textos até 10 páginas, apenas alguns segundos são necessários para o Sobek retornar os resultados do processo de extração de conceitos.



**DICA:** Se você não estiver satisfeito com os resultados, você pode alterar o tamanho dos gráficos que retornam, alterando as Configurações no menu superior.

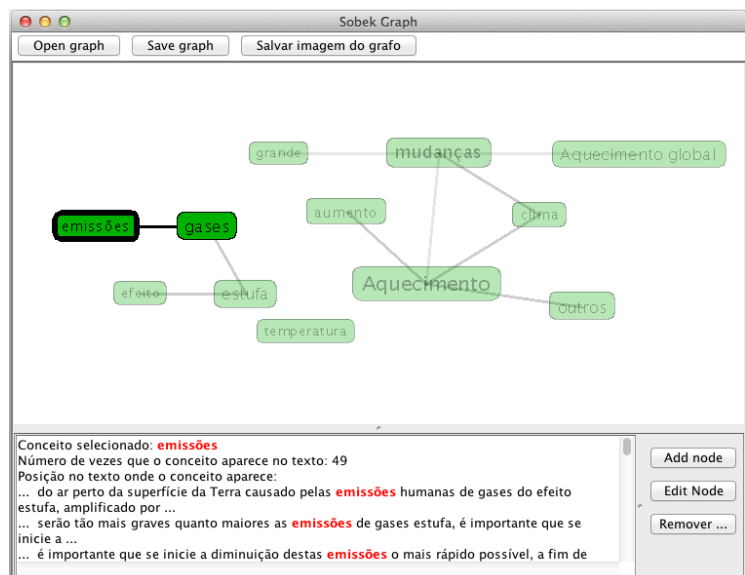
## 4. Visualizar e manipular gráficos

Na nova janela aberta, para exibir os resultados, você pode clicar e arrastar os nodos para ajustar o gráfico na aparência que você preferir. Observe que nodos maiores representam termos mais frequentes. Você também pode **adicionar**, **editar** e **remover** nodos, selecionando os botões correspondentes no lado direito da janela.



## 5. Explorando o gráfico

Observe também que quando você passa o mouse sobre um nodo só os seus termos conectados são realçados. Os outros são desativados. E ao clicar no nodo, algumas informações são apresentadas na parte inferior da janela, tais como o número de vezes que o termo correspondente aparece no texto e as sentenças em que aparecem.



## 6. Salvando os resultados

No fim, você pode querer salvar o grafo ou a sua imagem (JPG) para referência futura. Se você pretende usar o grafo na ferramenta Sobek futuramente, você deve usar a opção Salvar grafo. Para abri-lo mais tarde, você pode usar a opção de menu **Arquivo**, e então **Abrir grafo** a partir da janela principal do Sobek.



## E o que dizer do crocodilo?

Sobek foi uma divindade egípcia representada por um ser humano com uma cabeça de crocodilo.

Ele foi o deus que criou o rio Nilo, um símbolo de poder e força.

## Depoimentos de Alunos



"Para identificar os termos relevantes e analisar as relações destacadas no grafo eu tive que reler o texto várias vezes. Foi assim que eu fui capaz de contrastar o grafo com o texto."

Aluna do 2º ano do ensino médio, usando o Sobek em uma tarefa de resumo de texto.



"A possibilidade de extrair as principais informações a partir de um texto é muito útil, mesmo tendo que 'limpar' manualmente os termos [...]. Depois de observar se o gráfico fez sentido, a escrita do resumo foi fácil. Eu só tive que olhar para o grafo e seguir as relações, transformando a parte gráfica em texto."

Estudante de Engenharia Mecânica usando o Sobek para resumir um texto sobre contabilidade.

## Estratégias Pedagógicas para usar o Sobek

### Compreensão da Leitura

#### Leitura

- Peça aos alunos para lerem um determinado texto.

#### Reflexão

- Deixe os alunos usarem o Sobek para extraírem uma representação visual dos termos relevantes e das relações do texto;
- Peça aos alunos para refletirem sobre a pertinência dos termos e relações identificados pela ferramenta.

#### Discussão

- Proponha uma discussão com os alunos usando um projetor e mostre-lhes o gráfico extraído pelo Sobek.

### Fazendo resumos

#### Pré-escrita

- Peça aos alunos para lerem um determinado texto;
- Deixe os alunos usarem o Sobek para extraírem uma representação visual dos termos relevantes e das relações do texto;
- Peça aos alunos para revisarem os termos e relações identificados pela ferramenta e editarem o grafo de acordo com o que eles acreditam ser apropriado;
- Solicite que salvem seus grafos como uma imagem.

#### Escrita

- Usando o diagrama editado como um organizador gráfico, peça aos alunos que comecem a própria escrita do resumo. Durante esta etapa, eles podem ir e voltar ao texto original quantas vezes forem necessárias.

#### Pós-escrita

- Peça aos alunos para usarem o Sobek para analisarem seus próprios resumos.
- Peça-lhes para compararem seus grafos com o obtido a partir da análise do texto original. Esta comparação deve levá-los a refletir e talvez fazer os ajustes adequados aos seus resumos.





A Universidade Federal do Rio Grande do Sul - UFRGS é uma instituição de ensino centenária. Ela está localizada em Porto Alegre - a capital do estado do Rio Grande do Sul - e oferece programas acadêmicos de todas as áreas do conhecimento, para todos os níveis de ensino, que vai do básico à pós-graduação.

Mais de 30 mil pessoas circulam pela universidade, formando esta que é uma das principais instituições de ensino superior do país.

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

O Sobek foi desenvolvido pelo Grupo de Pesquisa GTech.Edu da Universidade Federal do Rio Grande do Sul (UFRGS). Por favor, envie-nos dúvidas ou comentários para: [info@sobek.ufrgs.br](mailto:info@sobek.ufrgs.br)



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



GTech.Edu

PPGIE - UFRGS

Av. Paulo Gama, 110 - prédio 12105 - 3º andar sala 332

90040-060 - Porto Alegre (RS) - Brasil

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

## Apoio

