

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA NA EDUCAÇÃO

DANIEL EPSTEIN

**USO DO MINERADOR DE TEXTOS SOBEK COMO FERRAMENTA DE  
APOIO À COMPREENSÃO TEXTUAL**

PORTO ALEGRE

SETEMBRO - 2017

DANIEL EPSTEIN

**USO DO MINERADOR DE TEXTOS SOBEK COMO FERRAMENTA DE  
APOIO À COMPREENSÃO TEXTUAL**

Tese de Doutorado apresentada ao  
Programa de Pós Graduação em Informática na  
Educação (PPGIE) da Universidade Federal do Rio  
Grande do Sul, como parte das exigências para obtenção  
do título de Doutor em Informática na Educação.

Orientador: Prof. Dr. Eliseo Berni Reategui

PORTO ALEGRE

SETEMBRO - 2017

CIP - Catalogação na Publicação

Epstein, Daniel  
USO DO MINERADOR DE TEXTOS SOBEK COMO FERRAMENTA  
DE APOIO À COMPREENSÃO TEXTUAL / Daniel Epstein. --  
2017.  
170 f.  
Orientador: Eliseo Berni Reategui.

Tese (Doutorado) -- Universidade Federal do Rio  
Grande do Sul, Centro de Estudos Interdisciplinares  
em Novas Tecnologias na Educação, Programa de Pós-  
Graduação em Informática na Educação, Porto Alegre, BR-  
RS, 2017.

1. Mineração de Texto. 2. Letramento. 3.  
Aprendizagem Significativa. 4. Compreensão Textual.  
5. Grafos. I. Reategui, Eliseo Berni, orient. II.  
Titulo.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os  
dados fornecidos pelo(a) autor(a).

## **AGRADECIMENTOS**

Primeiramente, gostaria de agradecer ao meu orientador, professor Eliseo Reategui, por todos os ensinamentos ao longo dos anos. Foram muitos anos de trabalho e aprendizado, antes mesmo de eu iniciar o doutorado, anos esses em que pude crescer de forma pessoa e profissional. O meu mais sincero agradecimento e admiração pela pessoa e profissional que tu és.

Agradeço também à minha família por todo apoio incondicional e educação dados ao longo da minha vida.

Agradeço à minha esposa Daniela por estar junto de mim nas longas noites em claro que passei e por tantas mais que virão junto com a Rafaela.

Agradeço por fim aos meus colegas de doutorado e meus amigos, em especial ao Otávio Acosta e a Simone Emer pela ajuda imensurável que me deram ao longo dos anos, sem nunca hesitar.

“I think I can solve the falling problem with a jetpack. Can you try to get me the parts?

That’s all I do, kiddo.”

Saturday Morning Breakfast Cereal

# SUMÁRIO

<b>1. Introdução.....</b>	<b>14</b>
<b>2. Compreensão Textual .....</b>	<b>21</b>
2.1 Etapas da leitura.....	22
2.2 Etapas da leitura de textos e compreensão textual.....	28
2.4 Considerações sobre o capítulo.....	32
<b>3. Teoria da Aprendizagem Significativa.....</b>	<b>33</b>
3.1 Considerações sobre o capítulo.....	39
<b>4. Mineração de Texto.....</b>	<b>41</b>
4.1 Conceitos de mineração de textos.....	41
4.2 Aplicações da mineração de textos .....	45
4.3 Minerador de textos Sobek .....	49
4.4 Considerações sobre o capítulo.....	54
<b>5. Metodologia.....</b>	<b>56</b>
5.1 Aprimoramento da ferramenta Sobek.....	56
5.1.1 Alteração da ferramenta de mineração de textos Sobek.....	57
5.1.2 Funcionamento do Sobek .....	60
5.1.3 Validação da ferramenta de mineração Sobek .....	65
5.2 O minerador de textos Sobek como apoio à compreensão textual .....	73
5.3 Avaliação da Ferramenta Sobek no Apoio à Compreensão Textual .....	77
5.3.1 Pesquisa preliminar .....	78
5.3.3 Estudo da fase final .....	91
5.3.2.1 Sujeitos da pesquisa .....	91
5.3.2.2 Materiais das intervenções .....	92
5.3.2.3 Descrição do experimento.....	93
<b>6. Resultados e discussão .....</b>	<b>96</b>
6.1 Análise dos dados coletados referentes ao estudo acerca da influência do minerador de textos Sobek na compreensão textual dos estudantes.....	96
6.2 Análise da ação dos usuários sobre o grafo .....	102
6.3 Questionário com estudantes participantes da pesquisa .....	114
<b>7. Conclusão.....</b>	<b>120</b>
<b>Referências .....</b>	<b>126</b>

<b>Apêndice A .....</b>	<b>135</b>
<b>Apêndice B .....</b>	<b>137</b>
<b>Apêndice C .....</b>	<b>139</b>
<b>Apêndice D .....</b>	<b>142</b>
<b>Apêndice E .....</b>	<b>143</b>
<b>Apêndice F.....</b>	<b>145</b>
<b>Apêndice G.....</b>	<b>147</b>
<b>Apêndice H.....</b>	<b>155</b>
<b>Apêndice I.....</b>	<b>161</b>
<b>Apêndice J.....</b>	<b>167</b>

## LISTA DE QUADROS

Quadro 1: Lista de termos considerados relevantes pelos especialistas para o <i>primeiro</i> artigo .....	67
Quadro 2: Lista de termos que compõe a lista gold standard, que são considerados relevantes pelo Sobek e se ambas listas concordam com a classificação (para o <i>primeiro</i> artigo). .....	68
Quadro 3: Lista de termos considerados relevantes pelos especialistas para o <i>segundo</i> artigo .....	69
Quadro 4: Lista de termos que compõe a lista gold standard, que são considerados relevantes pelo Sobek e se ambas listas concordam com a classificação (para o <i>segundo</i> artigo). .....	70
Quadro 5: Respostas para cada questão do texto 1 e número de alunos que responderam corretamente sem apoio do Sobek. ....	97
Quadro 6: Respostas para cada questão do texto 2 e número de alunos que responderam corretamente sem apoio do Sobek. ....	97
Quadro 7: Respostas para cada questão do texto 1 e número de alunos que responderam corretamente com apoio do Sobek.. ....	98
Quadro 8: Respostas para cada questão do texto 2 e número de alunos que responderam corretamente com apoio do Sobek.. ....	98
Quadro 9: Respostas para cada questão do texto 1 e número de alunos que responderam corretamente sem apoio do Sobek.. ....	99
Quadro 10: Respostas para cada questão do texto 2 e número de alunos que responderam corretamente sem apoio do Sobek. ....	99
Quadro 11: Respostas para cada questão do texto 1 e número de alunos que responderam corretamente com apoio do Sobek... ..	100
Quadro 12: Respostas para cada questão do texto 2 e número de alunos que responderam corretamente com apoio do Sobek.. ....	100

## LISTA DE FIGURAS

Figura 1 Processo de mineração de textos. Figura adaptada de Tan (1999) .....	42
Figura 2 Interface de uso do Sobek. Retirado de Lorenzatti (2007) .....	50
Figura 3. Interface do Sobek com texto. Retirado de Lorenzatti (2007) .....	51
Figura 4 Representação da base de termos extraída pelo Sobek, com os termos e seus relacionamentos. Retirado de Lorenzatti (2007) .....	53
Figura 5 Representações de grafos produzidos pelo Sobek. À esquerda, imagem retirada de Macedo (2010); à direita, imagem retirada de Lorenzatti (2007). .....	53
Figura 6 Interface original do minerador Sobek a esquerda (Lorenzatti, 2007) e interface remodelagem apresentada a direita. ....	57
Figura 7 Visualização do grafo utilizando <i>Interactive Graph Drawing API</i> após alterações iniciais. ....	59
Figura 8 Visualização do grafo resultante da operação de extração de termos do Sobek. É utilizado a <i>Prefuse API</i> . ....	59
Figura 9 Apresentação das informações do termo “PPGIE” .....	60
Figura 10 Imagem da praia de Ipanema em Porto Alegre.. ....	79
Figura 11 Jovem andando de skate pelo “calçadão” da praia de Ipanema, Porto Alegre. ....	80
Figura 12 Gráfico referente ao índice de satisfação com o minerador Sobek e com as atividades desenvolvidas. ....	85
Figura 13 Nota dos alunos para a pergunta se o grafo do Sobek representa bem o texto. ....	87
Figura 14 Imagem contendo o JSON produzido pelo Sobek a partir da interação de um estudante com o grafo. ....	104
Figura 15 Representação visual da Figura 14. ....	104
Figura 16 Grafo formado pelo Sobek a partir do texto "O dono da bola" de Ruth Rocha .....	107
Figura 17 Grafo referente ao texto “O dono da bola” e alterado pelo estudante A. ....	108
Figura 18 Grafos construídos pelos estudantes de primeiro ciclo para o texto "O dono da bola". ....	109
Figura 19 Grafo construído pelo Sobek a partir do texto de Luis Fernando Veríssimo .....	110
Figura 20 Grafos construídos pelos estudantes de terceiro ciclo para o texto "Defenestrar" .....	111

Figura 21 Grafo produzido pelo estudante B, estudante com melhor desempenho no terceiro ciclo .....	111
Figura 22 Grafo produzido pelo estudante C, estudante com menor número de acertos no texto de Luis Fernando Veríssimo. ....	112
Figura 23 Respostas dos estudantes a questão 1 do questionário de avaliação da atividade. ....	115
Figura 24 Resultados obtidos das respostas das questões 2 e 3 .....	116
Figura 25 Gráfico representando a dificuldade dos estudantes no uso da ferramenta Sobek .....	117

## LISTA DE SIGLAS

ANA - Avaliação Nacional da Alfabetização

EaD - Educação à distância

MT - Mineração de texto

PISA - *Programme for International Student Assessment* (Programa de Avaliação Internacional de Estudantes)

PPGIE - Programa de Pós-Graduação em Informática na Educação.

UFRGS - Universidade Federal do Rio Grande do Sul

VWFA - *Visual word form area* (área de formação visual das palavras)

## RESUMO

A presente tese tem por objetivo investigar os efeitos do uso do minerador de textos Sobek no processo de leitura e compreensão textual de estudantes. Este minerador de textos é capaz de extrair informações relevantes de textos e representá-las de forma gráfica. Esta tese está apoiada nas teorias de *aprendizagem significativa*, de uso de mapas conceituais para representação de conhecimento e em pesquisas que apontam que representações gráficas de palavras auxiliam na leitura de textos e na sua decodificação. De acordo com a pesquisa de David Ausubel, a aprendizagem significativa ocorre através da *assimilação* de novos conceitos e ideias e associação destas ao conhecimento que a pessoa já possui. Através da utilização de um minerador de textos com representação gráfica de informações, busca-se apresentar aos estudantes uma representação visual de textos. Esta representação se assemelha a de um mapa conceitual, de forma a auxiliar no processo de compreensão e assimilação de informações pelos estudantes. Nesta representação, ligações entre termos considerados relevantes pelo minerador auxiliam no entendimento destes termos e simbolizam relações presentes no texto, fato esse que pode auxiliar os estudantes a compreenderem melhor o texto e relacionarem novas informações àquelas que já possuem.

Nesta tese, foi realizado um estudo para auxiliar estudantes nas atividades relacionadas ao letramento. A pesquisa se caracteriza como mista (qualitativa e quantitativa). A coleta de dados se deu a partir da aplicação de questionários com professores e alunos, além de avaliações com o objetivo de verificar contribuições do uso da ferramenta a partir de seu uso do ponto de vista do letramento. Como resultado, encontramos que estudantes que utilizaram o Sobek obtiveram um número mais elevado de respostas corretas nas atividades de interpretação de textos. Em média, os alunos acertaram 66% das questões quando utilizando o minerador de textos Sobek, contra apenas 47% das questões que eram respondidas sem o apoio do minerador. Outro resultado apresentado é o alto grau de satisfação de alunos e professores quanto à tecnologia e seu uso em sala de aula. Além destes resultados, obtivemos uma avaliação acerca da capacidade do minerador de textos de extrair termos considerados relevantes ao texto.

**Palavras-chave:** Compreensão textual, Sobek, Leitura, Aprendizagem significativa, Mineração de textos, Grafos, Letramento

## ABSTRACT

This thesis aimed to investigate the effects of using Sobek Text Miner to improve literacy. Sobek is a tool capable of extracting relevant information from texts and representing them in a graphical way. The thesis is supported by *meaningful learning* theory, conceptual maps theory and several research theories which indicate that graphical representation of words may improve reading capability and word decoding. According to David Ausubel, meaningful learning occurs through the *assimilation* of new concepts and ideas and association of those to what the person already knows. Using text mining with graphical representation of information, we seek to provide students with a graphical representation of a text. This text representation is similar to a concept map, helping students assimilate and comprehend that information. In Sobek's representation, the relationship between terms considered relevant to text comprehension may assist students to better understand the meaning of each term and demonstrate relationships that are presented in the text, improving context comprehension. Furthermore, the relationship between terms may help information assimilation, once it relates the new information with previous known information.

This project conducted a study using Sobek text miner in classroom to support student's literacy. In order to assess the tool's possible benefits in reading and comprehension activities, we designed a series of classroom activities. To evaluate those activities, qualitative and quantitative approaches were used.

The study was conducted in two primary schools, with students from 5<sup>th</sup> grade and 8<sup>th</sup> grade. Interviews were also made with the teachers and students, inquiring them about the tool's and main functions and its ability to help students from a literacy point of view. The study shows that students answered more correct question when using Sobek than when no support technology was used. Also, both students and teachers approved the software and agreed that it does improve student's text comprehension. It also describes an evaluation of Sobek's capability to extract terms considered relevant for text comprehension.

**Keywords:** Text Comprehension, Sobek, Reading, Meaningful learning, Text mining, Graphs, Literacy

# 1.Introdução

Historicamente, um indivíduo é considerado alfabetizado quando é capaz de ler e escrever uma breve sentença sobre a própria vida, ou consegue ler e compreender em linhas gerais uma notícia de jornal. No entanto, as expectativas de hoje em dia com relação à leitura e escrita envolvem habilidades que vão muito além destas mais simples. Warschauer (2006) define letramento como a capacidade de dar significado aos textos com base na própria experiência, a habilidade de usá-los de maneira funcional e a capacidade de analisá-los e transformá-los de maneira crítica. Nesta visão mais contemporânea, espera-se que os indivíduos sejam capazes de utilizar a linguagem oral e escrita demonstrando, sua compreensão do mundo, sendo capazes de se comunicar, de participar na resolução de problemas e na tomada de decisões (Jenner, 2003).

Vivemos hoje imersos em um mundo onde precisamos necessariamente aprender a nos comunicar através da linguagem escrita (Sticht et al., 1974). Vieira (1981) constata que existe uma estreita relação entre a falta de leitura e a dificuldade de escrita. Este problema, segundo Marcuschi e Xavier (2005), está associado à linguagem que o indivíduo utiliza no seu dia-a-dia, linguagem essa baseada no código mais restrito da fala, que não contempla as construções mais complexas e elaboradas necessárias na produção textual.

As dificuldades apresentadas pelos estudantes na compreensão de texto e na produção escrita são problemas bastante sérios que vêm sendo discutidos tanto no Brasil quanto em outras partes do mundo. Dados do Ministério da Educação e Cultura (MEC) apontam que, já no ensino fundamental, no que se refere ao mau desempenho escolar, a questão da leitura e da escrita tem sido um dos principais problemas. Tais problemas acompanham os estudantes em toda sua trajetória escolar, afetando também seu desempenho no ensino superior. As escolas brasileiras têm ocupado as últimas posições nas avaliações PISA, colocando o país nas últimas posições dentre os países que participam da avaliação (OECD, 2013). Tais dados demonstram a necessidade de desenvolver métodos e ferramentas que possam apoiar os estudantes em atividades envolvendo a produção textual, atividades estas presentes em grande parte das disciplinas oferecidas em nível superior.

Os problemas enfrentados por diversos estudantes na compreensão de textos trazem consequências para diversas áreas da vida deles. Não é apenas uma questão de saber ler; é preciso entender o que se está lendo. Sem a compreensão, a leitura torna-se um processo sem finalidade e o aprendizado da leitura torna-se um processo que não será utilizado pelo aluno. A questão de compreensão textual vai além do simples aprendizado de fonemas e gramática. Dificuldade em leitura, escrita e matemática também podem reduzir o interesse dos estudantes em outras disciplinas, sendo fortemente relacionadas a outros problemas no desenvolvimento acadêmico deles (Korhonen et al 2014). De acordo com Dirks e colegas (2008), a dificuldade de leitura está constantemente associada à dificuldade em outras áreas acadêmicas, tais como matemática.

Aproximadamente 75% dos estudantes com alguma dificuldade em aprendizagem manifestam dificuldade social que não estão presentes em estudantes que não possuem tais dificuldades (Sridhar & Vaughn, 2001). Muitas vezes, as dificuldades de aprendizagem dos estudantes estão associadas com dificuldades no *letramento* (do inglês, *literacy*). De acordo com Weinstein e Siever (2003), estudantes com problemas de aprendizado precisam de constante apoio e dispõem de mais energia para realizar tarefas consideradas ordinárias do que aqueles alunos sem problemas de aprendizado. Isso afeta a interação dessas pessoas com as demais e dificulta o processo de aprendizagem. Weinstein e Siever também enfatizam que as dificuldades de aprendizagem são características individuais e que podem ser minimizadas por um processo educacional adequado.

É comum estudantes com dificuldades de leitura terem dificuldades na compreensão e na organização das informações contidas nos textos. Diversos pesquisadores propõem que o uso de tecnologias associadas à leitura podem amenizar tais problemas e auxiliar no processo de compreensão textual. Nandhini and Balasundaram (2013) propuseram uma aplicação capaz de extrair sumários de textos, auxiliando os estudantes no processo de organização de informações. Os experimentos realizados por eles demonstram que tal método tem efeitos positivos quando utilizado junto com textos educacionais. Wei et al. (2012) propuseram uma abordagem diferente, direcionada à integração de diferentes técnicas de leitura em e-books a fim de demonstrar como a compreensão textual é ampliada pelo uso destas técnicas. Calvo et al. (2011) propuseram um editor de texto que incentiva a reflexão de textos através do

questionamento ao estudante sobre o conteúdo do mesmo, sua estrutura e coerência. Já Villalon e colegas (2008) usaram um editor de textos para formular perguntas relacionadas ao texto, buscando incentivar o pensamento crítico dos estudantes e aprimorar sua compreensão do texto lido. Tais pesquisas realizadas e os seus resultados positivos demonstram a importância de se investigar como a tecnologia pode contribuir para atividades relacionadas ao letramento.

Outra abordagem de pesquisa na área vem demonstrando que o uso de tecnologias na forma de organizadores gráficos em diferentes atividades de aprendizagem tem obtido resultados benéficos para os estudantes. Isso é especialmente enfatizado em atividades que envolvem compreensão de padrões, tais como informações temporais, relatos de causa/consequência, informações descritivas, generalização, etc. (Marzano et al. 2001). Do ponto de vista de atividades de escrita, o uso de organizadores gráficos tem auxiliado os estudantes a escrever seus textos, ajudando-os a melhor organizar sua escrita e suas ideias (Beissner et al. 1994). Nesta linha de pesquisa, Novak e Cañas (2006) propuseram um modelo computacional para auxiliar no processo de organização de conceitos. Este modelo ficou conhecido como “*mapa conceitual*” (do inglês, *concept map*) e permite representar proposições acerca de eventos e objetos. Um mapa conceitual pode ser visto como um organizador gráfico no qual os seus nodos representam conceitos e a conexão entre tais nodos representam a relação entre esses conceitos (Nesbit and Adesope 2006).

A teoria e uso de mapas conceituais foi desenvolvida em 1972, durante um curso ministrado por Novak onde ele buscava entender as mudanças que ocorriam nas crianças durante o aprendizado de ciências (Novak & Musonda, 1991). As ideias de Novak eram fortemente baseadas na Teoria de Assimilação de David Ausubel (1963). A base da teoria de aprendizagem significativa de Ausubel é que o aprendizado ocorre ao se assimilar novos conceitos e proposições no corpo de conhecimento que a pessoa já possui.

Alguns autores propuseram diferentes métodos de conexão entre conceitos relevantes do texto e de visualização da informação. Reader e Hammond (1994) conduziram experimentos nos quais usaram técnicas de hipertexto para auxiliar os estudantes a construir representações de conhecimento. Os resultados destes estudos demonstraram que os indivíduos que utilizaram tal técnica tiveram melhor desempenho que aqueles que não utilizaram. Chang et. al. (2001) desenvolveram um processo

computacional que permite a criação de representação gráfica de conceitos através de uma abordagem de “construção partindo de uma estrutura”.

Uma abordagem diferente para o uso de organizadores gráficos voltados à compreensão textual tem como foco o incentivo aos estudantes a fazerem um *brainstorm* sobre informações conhecidas por eles acerca de um determinado tópico, representando estas informações graficamente e conectando os diferentes conteúdos. Essa visualização gráfica das ideias e relações permite ao estudante ter uma representação mais concreta das informações que são mais relevantes para um determinado assunto. De acordo com Fisher et al (2002), organizadores gráficos são uma das estratégias de letramento com impacto mais significativo na aprendizagem e na compreensão de conteúdos.

Organizadores gráficos (em especial mapas conceituais) também têm um papel importante para sumarização de textos. De acordo com Chang et al. (2001), a organização dos conceitos centrais do texto e a ligação entre elas permite ao estudante uma ênfase maior nas ideias principais que devem ser descritas. Brown e Day (1983) afirmam que há uma forte relação entre a organização de ideias e conceitos e o processo de sumarização de textos.

Alguns estudos têm como foco a identificação das diferenças que existem entre “bons leitores” (com alto grau de compreensão textual e facilidade de leitura) e “maus leitores”. Entre essas características, podem-se destacar diversos fatores, tais como: capacidade de reconhecer informações que se deseja extrair de textos, visão geral da estrutura do texto, associação de ideias e informações, capacidade de prever o desenrolar do texto, revisão de informações conflitantes (sejam informações contidas no texto ou aquelas obtidas previamente pelo leitor) e reconhecimento de vocabulários e acentuação. Estudantes jovens ou com pouca prática de leitura demonstram dificuldade na realização dos processos acima mencionados (Pressley & Afflerbach, 1995, Córdón & Day, 1996). Porém, é possível estimular maus leitores a se tornarem bons leitores através de abordagens que instigam o leitor a refletir sobre o texto. Entre essas abordagens destacam-se o questionamento sobre o texto e sua ideia principal, a construção de imagens mentais do texto, a sumarização do texto e a análise da relação dos personagens e fatos ocorridos no texto (Block, & Pressley, 2002; Moats, 1998). A maioria dos leitores com os quais foram trabalhadas estratégias de compreensão textual são mais propensos a aprender, a desenvolver e usar essas estratégias de forma espontânea (Block, & Pressley, 2002). Pesquisas demonstram que reforçar a ortografia

das palavras na forma escrita é uma maneira de auxiliar leitores com dificuldade de compreensão textual. Quando um leitor reconhece uma palavra, ele associa sua ortografia a sua fonologia e isso auxilia no seu processo de compreensão da palavra. As pesquisas de Rosenthal e Ehri (2008) e de Ehri e Wilce (1979) demonstraram que a visualização de novas palavras na forma escrita melhorava a memorização de crianças tanto do ponto de vista de vocabulário oral quanto do significado destas palavras. Nos seus experimentos, estudantes que visualizavam a ortografia de palavras desconhecidas obtiveram melhores resultados tanto ao responderem perguntas acerca de textos que continham essas palavras quanto na utilização de tais palavras em textos escritos pelos próprios estudantes.

As pesquisas acerca do uso de tecnologias para auxiliar no letramento e seus resultados positivos fundamentam a proposta de utilização da técnica de mineração e representação gráfica de um texto com o propósito de auxiliar a compreensão textual. Esta tese foi baseada no aprimoramento de um minerador de textos chamado Sobek para apoiar atividades de leitura e compreensão textual. O Sobek é uma ferramenta desenvolvida fundamentalmente para fins pedagógicos, capaz de extrair um conjunto de termos e relacionamentos a partir da análise léxica e estatística de um texto. Apesar de sua simples forma de utilização, a ferramenta possui uma ampla gama de configurações, possibilitando sua personalização para diferentes fins e ambientes.

O minerador Sobek extrai as informações consideradas por ele mais relevantes em um texto e as representa na forma de um grafo no qual os termos considerados importantes são os nodos e as ligações entre os nodos representam as relações dos termos no texto. Esta representação gráfica se aproxima de um mapa de conceitos e teve sua construção baseada na teoria da aprendizagem significativa de David Ausubel. Além da representação gráfica do texto, é permitido ao usuário manipular nodos e relações do grafo, fomentando assim o pensamento crítico e a reflexão acerca do texto. A representação de nodos contendo a ortografia dos termos juntamente com representação não textual busca auxiliar na compreensão de textos do estudante e na organização de ideias.

Resultados acerca da influência do minerador Sobek e de sua capacidade de extrair informações de textos são apresentados ao longo desta tese. Estes resultados,

aliados aos problemas mencionados sobre a dificuldade de leitura dos estudantes, levaram-nos a definir a seguinte questão de pesquisa:

*De que maneira a ferramenta de mineração de textos Sobek pode apoiar o processo de leitura e compreensão textual?*

A partir desta questão, foram elencados os seguintes objetivos:

**Objetivo Geral: “Investigar como a ferramenta de mineração de textos Sobek pode apoiar o processo de leitura e compreensão textual”.**

**Objetivos Específicos:**

- Aprimorar e avaliar a ferramenta Sobek no que diz respeito à sua capacidade de extração de termos e como elemento de apoio à leitura e compreensão textual;
- Elaborar práticas pedagógicas que podem ser utilizadas associadas ao Sobek a fim de potencializar sua utilização para apoio aos processos de leitura e compreensão textual.

Além de contribuir com o avanço das pesquisas sobre o emprego da tecnologia para apoio ao letramento, os resultados desta tese devem prover ferramentas e práticas que contribuam para minimizar problemas de leitura e compreensão textual enfrentado por leitores iniciantes.

A presente tese está estruturada da seguinte forma:

- O capítulo 2 apresenta uma revisão bibliográfica acerca dos problemas relacionados à compreensão textual e suas implicações para o desenvolvimento dos estudantes, além de pesquisas que buscam solucionar tais problemas e suas fundamentações teóricas;
- O capítulo 3 descreve técnicas de mineração de texto e representação de informação, bem como o minerador de textos Sobek;
- No capítulo 4 é apresentada a metodologia de pesquisa, as ferramentas que serão utilizadas, seus desenvolvimentos e resultados preliminares.
- No capítulo 5 apresentamos os resultados da pesquisa, bem como uma discussão acerca deles.
- O capítulo 6 apresenta a conclusão do trabalho e perspectivas de trabalhos futuros.



## 2. Compreensão Textual

Compreensão de textos pode ser entendida como um processo hierárquico de ações para identificação de letras, palavras e frases para dar um significado ao que está sendo lido (Tankersley, 2003). A leitura em si corresponde a apenas parte desse processo, que envolve ações como decodificação de símbolos, identificação de fonemas e associações semânticas. Compreensão textual requer, além de boas práticas de leitura, a compreensão daquilo que se está lendo, dando um significado à coleção de letras e símbolos e permitindo ao leitor associar um significado a eles (Block & Pressley, 2002; Cordón & Day, 1996). Essas ações ocorrem antes, durante e depois da leitura. De acordo com Beers (2003), é necessário abordar de forma explícita e sistemática todos os aspectos da compreensão textual, pois se concentrar em apenas um deles não é o suficiente para garantir que o leitor desenvolva essa habilidade com sucesso.

Pressley (2000) define a compreensão textual como a parte principal do letramento (do inglês, *literacy*). Proficiência em leitura e escrita é fundamental para o sucesso de um indivíduo, seja esse sucesso no campo pessoal ou profissional. Estudantes que têm dificuldade na área de leitura têm risco de obter baixo rendimento escolar e incorrer em problemas sociais e de comportamento, o que também acaba por interferir no campo emocional destes estudantes (Wawryk-Epp, 2004). Sridhar e Vaughn (2001) apontam que aproximadamente 75% dos alunos com alguma dificuldade de aprendizagem manifestam dificuldade em interação social. Muitas vezes, estas dificuldades de aprendizagem estão associadas a dificuldades de letramento. De acordo com Weinstein e Siever (2003), estudantes com problemas de aprendizagem precisam de constante apoio e dispõem de mais energia para realizar tarefas não necessariamente complexas. Weinstein e Siever enfatizam, contudo, que as dificuldades de aprendizagem são características individuais e que podem ser minimizadas por um processo educacional adequado.

Alguns estudos tentam encontrar as diferenças entre bons leitores (com alto grau de compreensão textual e facilidade de leitura) e maus leitores. Entre essas características, pode-se destacar: capacidade de reconhecer informações em textos, visão geral da estrutura do texto, associação de ideias, capacidade de prever o desenrolar do texto, revisão de informações conflitantes (sejam informações contidas no texto ou aquelas obtidas previamente pelo leitor) e reconhecimento de vocabulários e

acentuação. Estudantes com pouca prática de leitura tendem a mostrar mais dificuldade nos processos listados (Pressley & Afflerbach, 1995). Porém, é possível estimular maus leitores a se tornarem bons leitores através de abordagens que instiguem o leitor a refletir sobre o texto. Dentre essas técnicas, destacam-se o questionamento sobre o texto e sua ideia principal, a construção de imagens mentais do texto, a sumarização do texto e análise da relação dos personagens e fatos descritos no texto (Block, & Pressley, 2001; Moats, 1998). A maioria dos leitores com quem foram trabalhadas estratégias de compreensão textual são mais propícios a aprender, desenvolver e usar essas estratégias de forma espontânea (Block, & Pressley, 2002).

## **2.1 Etapas da leitura**

Ao realizar o processo de leitura, duas tarefas são realizadas simultaneamente: decodificação e compreensão do texto (Coltheart, 2005). É fundamental que o leitor esteja apto a realizar ambas as tarefas, uma vez que elas são igualmente importantes para o aspecto de compreensão textual. O rápido reconhecimento de palavras e fluente compreensão da linguagem facilitam o processo de leitura enquanto a deficiência em um desses aspectos pode prejudicar o outro. Sendo assim, para entender melhor o processo de leitura é necessário compreender as tarefas a ele associadas.

A compreensão da linguagem se refere à atribuição de significado a uma determinada frase ou texto. Há diferentes tipos de linguagem (linguagem coloquial, linguagem formal, etc.) e diferentes formas de se utilizar essa linguagem. Adultos, por exemplo, utilizam uma linguagem geralmente mais elaborada que crianças. A linguagem informal, quando utilizada por crianças, é geralmente baseada no contexto no qual elas estão situadas. De acordo com Seidenberg e McClelland (1989) o significado e a compreensão da linguagem estão diretamente relacionados ao contexto onde ela aparece.

A compreensão da linguagem tem um papel importante não apenas para que um indivíduo possa compreender algo explicitamente dito (ou lido), mas também para que compreenda ideias colocadas de maneira implícita. Ao considerar o contexto no qual a linguagem aparece, o leitor pode ser capaz de identificar informações que não estão escritas de forma direta no texto. Esse contexto pode aparecer de diferentes formas no texto escrito (estilo de escrita, perspectiva, significado de palavras chave) e isso torna

muitas vezes o contexto difícil de ser identificado. Uma boa compreensão da linguagem permite ao leitor mais facilmente identificar o contexto no qual este texto está situado, levando-o a assim melhor compreender o que está sendo lido.

A segunda atividade considerada essencial para a leitura é a decodificação, na qual ocorre a decodificação de símbolos. Apesar da obviedade da afirmação a seguir, é importante comentá-la: sem conseguir decodificar os símbolos que formam as palavras e frases de um texto, um indivíduo não poderá compreendê-lo. Portanto, é necessário que o leitor seja capaz de decodificar os diferentes símbolos que fazem parte de um texto, tais como letras, pontuação, acentuação, números, etc.

De acordo com Regina (1992), já na pré-escola e ensino fundamental os estudantes adquirem conhecimentos específicos sobre textos e o que a linguagem escrita representa. Os autores mencionam que crianças na pré-escola já algum conhecimento sobre textos e sobre a concepção de que há uma mensagem contida neles. Além disso, elas já possuem a capacidade de compreender que símbolos e textos dispostos em locais específicos contêm um significado próprio (tal como o "M" amarelo do *McDonalds*). No primeiro ano do ensino fundamental, os estudantes já aprendem que a linguagem oral pode ser traduzida na linguagem escrita e vice-versa. Com esse conhecimento eles desenvolvem a noção de que textos aparecem em diferentes formatos e tamanhos e devem ser lidos em uma determinada ordem para que façam sentido. Durante o desenvolvimento do aluno nos primeiros anos do ensino fundamental o processo de leitura ganha significado e os estudantes já estão cientes que cada palavra ou sentença tem um valor sintático e semântico associados. Além disso, já empregam a leitura com diversos objetivos e usam uma variedade de estratégias para aprimorarem sua compreensão textual (mesmo que não estejam cientes de estarem utilizando tais estratégias).

Decodificar um texto portanto está relacionado à habilidade de reconhecer e processar informação escrita. Inicialmente, este processo é realizado pela decodificação de símbolos que têm um significado próprio como, por exemplo, uma marca ou um lugar (a palavra "*Coca-Cola*", quando escrita na sua logomarca, é considerada um dos símbolos mais reconhecidos no mundo e é compreendido mesmo por pessoas que não sabem ler as letras que formam essa palavra). É possível reconhecer palavras que são

constantemente observadas pelo leitor através do formato delas e não das letras que a compõem. Esse processo é bastante comum em crianças que estão aprendendo a ler.

Ao ler um texto estamos decifrando símbolos. Decifrar consiste de um processo sistemático e regular de converter uma informação em outra. Por exemplo, se cada letra fosse associada a um número, para decifrar uma palavra seria necessário realizar o processo inverso (converter os números em letras). Este processo corresponde apenas à conversão de símbolos (no caso, o alfabeto) em informação (o significado de cada letra). Poder decifrar uma palavra permite ler a palavra mesmo que ela seja desconhecida. A habilidade de corretamente decifrar palavras é importante para a decodificação de um texto e permite ao leitor corretamente ler uma palavra nova para ele. Cabe ressaltar que há uma arbitrariedade nos símbolos que correspondem de forma aleatória aos signos de uma linguagem.

A partir da capacidade de decifrar os símbolos que compõem a palavra podemos associar "som" para cada símbolo. Cada letra no alfabeto está associada a um ou mais fonemas; um fonema é a menor unidade da linguagem falada. Ao expressar os símbolos na forma de um som, o leitor começa a reforçar caminhos sinápticos que conectam a forma da palavra ao som dela. Isso reforça o processo de compreensão textual e auxilia o leitor a reconhecer quando uma palavra está fora de contexto ou mesmo quando a palavra é nova para ele. O cérebro também associa o som de uma palavra ao seu significado. A leitura da palavra repetidas vezes auxilia nesse processo de associação.

Esta relação entre a ortografia da palavra e seu som é considerada um dos pilares básicos da compreensão textual. É comum que a maioria das pessoas seja capaz de ouvir e distinguir os sons correspondes às palavras, podendo também entender a linguagem falada. Esta associação entre som e palavra também ocorre durante a leitura de um texto. De acordo com Coltheart (2005), experimentos de leitura em voz alta demonstram que quando uma palavra lida possui um som considerado diferente daquele que deveria ser obtido seguindo as regras gramaticais, a pessoa experimenta um breve instante de conflito no qual é mais difícil compreender o que foi lido.

Ter fluência na leitura e um bom nível de compreensão está diretamente relacionado ao domínio acerca dos fonemas das palavras (Adams et. al., 1998). A forma como a decodificação de letras, *grafema* (uma ou mais letras que constituem um único fonema) e palavras ocorre ainda é discutido pelos pesquisadores mas é comum associar

de forma direta a compreensão textual ao processo de reconhecer a ortografia da palavra e ser capaz de identificar seu fonema (Baron, 1977; Seidenberg e McClelland, 1989; Coltheart, 2005).

Uma importante parte da decodificação de palavras envolve o vocabulário e o conhecimento prévio que a pessoa possui. O aspecto relacionado ao conhecimento prévio do indivíduo tem papel importante porque ao ler um texto a pessoa relaciona as informações lidas àquelas que já possuem. Para um leitor que nunca ouviu falar de futebol, por exemplo, frases como “*gol do Pelé*” possuem um significado diferente daquele dado por pessoas familiarizadas com o esporte. O leitor relaciona as experiências de vida dele com as informações lidas, fornecendo um significado próprio ao texto.

Estudos demonstram que alunos que possuíam conhecimento prévio sobre um determinado assunto apresentaram melhores resultados em testes de compreensão textual que os demais alunos quando os textos versavam sobre esses assuntos (Anderson & Pearson, 1984). Apesar disso, McKoon & Ratcliff (1992) argumentam que nem sempre um leitor consegue relacionar um texto com conhecimentos prévios adquiridos. Isso é especialmente recorrente quando uma pessoa lê uma informação nova, mesmo que seja sobre um assunto do qual ela possua conhecimento.

Para auxiliar no relacionamento da informação lida com conhecimentos prévios do leitor, é importante incentivá-los a se questionarem sobre os fatos do texto e porquê eles fazem sentido (Pressley et. al., 1992). Ao se indagar sobre alguma informação lida, o leitor consegue buscar argumentos que relacionam esse conteúdo com o conhecimento prévio que possui. O ato de pensar sobre o assunto e buscar informações não contidas no texto para responder às indagações auxilia na compreensão do que foi lido. Isso permite que uma relação seja criada entre a informação nova recém lida e aquelas sobre as quais o leitor tem conhecimento. Quanto maior o conhecimento e experiências do leitor, maior será a capacidade de relacionar as informações novas às que ele já possui. O vocabulário possui uma grande influência nesse aspecto, permitindo que o leitor possa entender o que está lendo.

Um leitor experiente possui um vocabulário mais vasto e um maior conhecimento de palavras. De acordo com LaBerge and Samuels (1974), saber ler a palavra corretamente (e, pronunciá-la corretamente) não garante que a pessoa entenda o

que está lendo. De acordo com os autores, isso decorre do fato que a leitura de uma palavra desconhecida requerer um esforço cognitivo muito grande, deixando a parte de reconhecimento e compreensão da palavra em segundo plano. Principalmente na educação de crianças, é possível que estas atentem em demasia na pronúncia correta da palavra e não deem atenção ao seu significado junto ao texto. Isso levou os autores a concluir que é importante não apenas ensinar as pessoas a ler mas também ensinar vocabulário e reconhecimento de palavras. Leitores habituais (fluentes) conseguem decodificar as palavras rapidamente, deixando assim a maior parte de sua atividade cognitiva para a compreensão do que se está lendo.

Tan and Nicholson (1997) realizaram um estudo para medir o impacto do reconhecimento de palavras na compreensão textual. Para o experimento, os autores dividiram uma turma em dois grupos. No primeiro, 10 palavras foram ensinadas focando no aspecto de reconhecimento de palavras (os estudantes leram as palavras até o momento em que conseguiam reconhecer imediatamente cada uma delas). O segundo grupo recebeu as mesmas 10 palavras, mas o enfoque foi no significado dessas palavras e sua semântica. Após essa primeira etapa, os estudantes deveriam ler um pequeno texto que continha essas palavras e responder questões sobre compreensão textual. Os estudantes que aprenderam a reconhecer as palavras responderam mais questões corretamente que aqueles que aprenderam apenas o seu significado. Esse estudo foi comprovado posteriormente por Breznitz (1997), demonstrando assim a importância do reconhecimento de vocabulário na compreensão textual. O reconhecimento de palavras não está apenas no reforço de leitura, mas também na importância em se enfatizar os fonemas que compõem a palavra (Snow et. al., 1998).

De acordo com Anderson e Freebody (1991), estudantes com boa compreensão textual geralmente têm um vocabulário maior que estudantes com má compreensão. Isso não indica que estudantes com vocabulários mais extensos tenham melhor compreensão textual, apesar de que alguns estudos apontem para essa direção. Um exemplo disso são os experimentos realizados por Beck e colegas (1982), nos quais um grupo de 104 palavras foi ensinado para alguns estudantes durante um período de tempo. Quando testes foram realizados com os estudantes, aqueles para quem o grupo de palavras foi ensinado obtiveram melhores resultados que os demais estudantes. Isso aponta para indicadores que consideram o ensino de vocabulário para crianças como sendo uma abordagem eficiente para melhorar o processo de compreensão textual.

De acordo com Regina (1992), os estudantes devem não apenas ser expostos a novos vocabulários, mas também devem aprender estratégias de leitura que lhes permitam reconhecer e compreender novas palavras à medida que elas aparecem no texto. Dentre as instruções que permitem ao leitor aprender um novo vocabulário estão:

- Leitores devem ser apresentados ao vocabulário em diferentes contextos e frases;
- Novos vocabulários devem ser associados com conhecimentos prévios;
- Deve haver relação entre os vocabulários novos e os já conhecidos;
- Leitores devem aprender a utilizar dicionários e outras fontes de referência durante a leitura;
- Leitores devem fazer uso desse novo vocabulário para que possam compreendê-lo de forma mais ampla e definitiva.

Além dessas instruções, Moats (1998) explica que a pessoa deve aprender a reconhecer a palavra de forma oral e escrita. Se a pessoa não souber pronunciar a palavra ela não será capaz de entendê-la ao vê-la escrita. O vocabulário oral é importante para o aprendizado e é crucial para a compreensão de texto.

Estudos recentes corroboram as teorias que dizem que as palavras já conhecidas pelo leitor não são lidas letra-a-letra (ou fonema-a-fonema), mas sim reconhecidas pelo seu formato e imediatamente decodificadas como sendo um conjunto de fonemas (Forster e Chambers, 1973). Para Baron e McKlipon (75), o processo de reconhecimento de palavras que já estão presentes no vocabulário do leitor é similar ao processo de compreensão da palavra ao ouvi-la. Assim, o processo de reconhecimento não passa necessariamente pela transformação da palavra em um conjunto de fonemas para que só então estes sejam interpretados. Isso significa que leitores que possuem um vocabulário mais extenso são capazes de associar mais rapidamente as palavras lidas ao seu significado, melhorando assim sua compreensão do texto.

A razão pela qual é possível identificar uma palavra já conhecida sem a necessidade de observar cada parte que a compõe está em uma área do cérebro chamada de área de formação visual das palavras (McCandliss *et. al.*, 2003). Esta área do cérebro é responsável não apenas pela leitura da palavra, mas também pela associação da palavra a outras áreas do cérebro, como a área responsável por dar um significado a um som que se está lendo.

## 2.2 Etapas da leitura de textos e compreensão textual

Através do avanço da tecnologia foi possível abordar as questões de letramento e compreensão textual investigando não apenas o comportamento dos leitores mas também a forma como estes processos ocorrem do ponto de vista cognitivo. Imagens neurológicas feitas em estudantes com e sem dificuldade de aprendizagem mostram que há diferenças sutis no funcionamento e estrutura cerebral desses dois grupos de indivíduos (Kibby & Hynd, 2001). Através dessas imagens, foram identificados alguns padrões neurais relacionados diretamente com o processo de leitura de textos. Esses mesmos padrões também estão relacionados com processos cognitivos de análise de palavras e fonemas, demonstrando assim que as dificuldades de leitura se estendem a questões além da decodificação de símbolos ou semântica.

Ler com fluência e compreender o que se está lendo são tarefas intrinsecamente ligadas ao reconhecimento de símbolos e letras e aos sons correspondentes a elas (Adams, 1994; Adams et. al., 1998). É importante que os estudantes que recém estão começando a ler sejam incentivados a conhecer e explorar textos contendo palavras curtas nos quais as letras representem de forma bem definida os sons correspondentes a elas. Textos longos, com palavras complexas ou muitas exceções de vocabulário adicionam um componente de estresse ao leitor, além de necessitarem de maior esforço cognitivo para decifrar as palavras, tornando a sua compreensão menos natural e mais difícil (Duncam e Seymor 2003). Palavras curtas são também mais facilmente memorizadas pelos leitores, permitindo que a memória visual auxilie no processo de leitura e compreensão textual. Como mencionado anteriormente, a área no cérebro responsável pela formação visual das palavras (VWFA) é responsável não apenas pela leitura da palavra, mas também por sua associação com outras, seu significado e seu som (McCandliss *et. al.*, 2003). A VWFA se desenvolve à medida que a criança começa a ter contato com textos escritos, possibilitando a identificação de símbolos que correspondem a letras e palavras (Maurer et. al., 2006). No primeiro momento, essa área do cérebro consegue de forma lenta e custosa transformar letras em fonemas. À medida que vai se desenvolvendo, ela consegue processar informações mais complexas e com maior velocidade, chegando ao estágio maduro de desenvolvimento no período da adolescência (Sandak et. al., 2004). Porém, alcançar este estágio não está necessariamente relacionado à capacidade

cognitiva do indivíduo, uma vez que esta habilidade é desenvolvida a partir da experiência e conhecimento da pessoa.

A VWFA não identifica uma determinada fonte, cor ou desenho da palavra, mas sim a estrutura ortográfica e suas letras (Dehaene et. al., 2004). Há diversos estudos que mostram que palavras bem estruturadas, frequentes e com ortografia simples são rapidamente identificadas por essa área do cérebro e podem ser mais facilmente compreendidas pelo leitor (Adams 2011). Cada letra da palavra é reconhecida pelo cérebro, que associa as imagens às palavras mais frequentemente observadas. À medida que as letras são lidas e as palavras são compostas, outras regiões do cérebro são ativadas para dar continuidade aos demais processos de caráter linguístico e semântico. Através desse processo a palavra é lida e reconhecida pelo leitor. A VWFA é responsável unicamente pelo reconhecimento das palavras e sua ortografia, não sendo esta ação relacionada com a linguagem ou interpretação da informação.

O cérebro envia a palavra reconhecida pela VWFA juntamente com sua ortografia para a área responsável pela interpretação fonológica da palavra (Adams 2011). Esta segunda área do cérebro reconhece cada letra da palavra e o fonema associado a ela. Uma vez que o 'som' da palavra tenha sido interpretado, uma série de outras áreas do cérebro são ativadas, sendo elas responsáveis por dar um significado à palavra. Diversos pesquisadores realizaram experimentos no sentido de comprovar que o mapeamento da ortografia das palavras para a sua fonologia é o processo central responsável pela capacidade de ver uma palavra escrita e entender o que ela significa (Adams, 1994; Dehaene, 2009).

Após reconhecer a palavra, um significado é associado a ela. Esse processo de compreender o significado de uma palavra envolve as experiências passadas do leitor e diversas outras áreas do cérebro. Palavras que se relacionem com movimentos das mãos, por exemplo, ativam além das áreas cognitivas relacionadas à leitura, áreas motoras responsáveis pelo movimento das mãos (Cramer, et. al. 2008). Outros pesquisadores demonstraram que textos relacionados à comida e ingestão de alimentos ativam áreas relacionadas à digestão (Olivetti et. al., 2009; Palmiero et. al., 2009).

Além destas áreas ativadas pelo significado das palavras, várias outras são ativadas com a intenção de fornecer informações relacionadas à palavra ao leitor. Algumas áreas do cérebro são responsáveis por identificar a relação que existe entre

duas palavras; também há áreas responsáveis por identificar as combinações de palavras que dariam sentido ao que foi lido e outras para indicar o significado da palavra em um determinado contexto (Rodd et. al., 2005). Todas essas áreas do cérebro trabalham juntas, sendo ativadas pela região responsável pela fonologia da palavra. E todas elas são responsáveis pelas funções relacionadas à linguagem, seja ela escrita, lida, falada ou ouvida.

A atividade neural decorrida do reconhecimento de uma palavra não é um fato isolado sem consequências futuras. Essa atividade ocorre através do envio de impulsos de ativação de uma área para a outra, que por sua vez responde a esse impulso. Essa relação entre as duas áreas do cérebro cria uma conexão entre áreas inicialmente isoladas, conexão essa que tem sua força aumentada a cada troca de impulsos. Como consequência, essas áreas tornam-se interligadas e sua ligação torna-se mais forte cada vez que um impulso é disparado entre elas (Hebb, 1949; Goldinger e Azuma, 2003). Esse fortalecimento de áreas do cérebro quando estimuladas faz com que pessoas que leiam com maior frequência e se tornem melhores leitores, reconhecendo e interpretando o texto mais rapidamente. Quanto mais for praticada a leitura, mais facilmente cada palavra será reconhecida pela VWFA e menos esforço será necessário para ler e compreender textos. Por outro lado, leitores com pouca experiência possuem ligações sinápticas mais fracas, tornando o processo de decodificação do texto mais dispendioso e o reconhecimento de palavras um processo mais lento.

Fracas ligações sinápticas podem resultar em baixa compreensão textual. Quando uma palavra é lida, o cérebro pode interpretar incorretamente a palavra, impedindo a compreensão do texto. Também é possível que a palavra possa ser mal interpretada ou mesmo confundida com algum homônimo. Esses empecilhos que surgem durante a leitura podem tornar o processo extremamente desgastante para leitores iniciais ou pouco qualificados, desestimulando esses leitores. Também pode ocorrer do processo de reconhecimento da palavra não apresentar conexões fortes o suficiente com a área responsável pela interpretação da palavra, fazendo com que o leitor leia mas não compreenda o que está sendo lido.

Uma forma de auxiliar leitores com dificuldade de compreensão textual é reforçar a ortografia das palavras na forma escrita. Quando um leitor reconhece uma palavra, ele associa a ortografia dela a sua fonologia. Isso forma conexões sinápticas

que são reforçadas à medida que a palavra encontrada é relida. É comum aprender a decodificar palavras através do uso do vocabulário oral e por isso reforçar a grafia junto com a fonologia da palavra permite que as ligações sinápticas se fortaleçam e que a compreensão textual seja facilitada. Ehri e Wilce (1979) e Rosenthal e Ehri (2008) demonstraram que a visualização de novas palavras de forma escrita melhorava a memória de crianças tanto do ponto de vista de vocabulário oral quanto do significado destas palavras.

Nestes experimentos, foi pedido às crianças de segundo ano e de quinto ano do ensino fundamental que aprendessem dois conjuntos de palavras usadas com pouca frequência, sendo que todas as palavras tinham uma representação visual associada. As listas dos estudantes do segundo ano tinham palavras que consistiam de consoante, vogal e consoante; as listas dos estudantes de quinto ano tinham palavras de duas a três sílabas. O conjunto de palavras fornecido tinha imagens, definição e um grupo de sentenças no qual a palavra aparecia. Além disso, os estudantes eram incentivados a usarem as palavras e fazerem perguntas sobre elas durante o tempo de duração do estudo. Os estudantes foram divididos em grupos e receberam cartões com a imagem das palavras. Um dos grupos recebeu a palavra escrita embaixo da imagem, enquanto o outro grupo recebeu apenas a imagem referente à palavra. O professor lia em voz alta cada palavra e mostrava a imagem correspondente aos alunos. Mesmo que a palavra estivesse escrita junto à imagem os alunos não precisariam lê-la (já que o professor mostrava a imagem) e nem foi pedido aos estudantes que lessem o que estava escrito. Mesmo assim, o estudo mostrou que o mero fato da palavra estar presente no cartão de forma escrita auxiliou na sua memorização. O simples fato de associar a pronúncia à grafia fez com que o grupo de alunos que continha a palavra escrita a aprendesse mais rapidamente que o outro grupo. Esse resultado foi percebido nos dois experimentos, tanto para os estudantes de segundo ano quanto para os estudantes de quinto ano. Além disso, um pós-teste foi realizado com os estudantes do quinto ano. Foi pedido que eles usassem as palavras que aprenderam em novas frases. O resultado foi que o grupo que recebeu as palavras escritas as utilizou com 50% a mais de frequência que o outro grupo.

## **2.4 Considerações sobre o capítulo**

Neste capítulo foram abordados conceitos relacionados à leitura, compreensão textual e aprendizagem significativa, relacionando tais conceitos ao letramento. Foram discutidas as diferentes etapas da leitura e como o processo de decodificação de símbolos ocorre durante estas etapas. A compreensão textual e as suas características também foram debatidos.

É enfatizado que durante as pesquisas de Wawryk-Epp (2004) e de Sridhar e Vaughn (2001) é traçada um paralelo entre um mau desempenho escolar e a capacidade dos estudantes relacionada a questões do letramento. Essas dificuldades são recorrentes ao longo da vida social e acadêmica do indivíduo, demonstrando a necessidade de se endereçar o reforço do letramento ainda nas fases iniciais da alfabetização. Esse reforço visa reduzir o impacto que a baixa compreensão textual tem sobre estudantes e aprimorar a compreensão de pessoas consideradas “más leitores”.

Por fim, é descrito o processo de reconhecimento visual de palavras e o papel de tal habilidade no auxílio à leitura. Essa característica neurológica do indivíduo ajuda na compreensão sintática das palavras lidas e deve reduzir o esforço cognitivo necessário para ler textos, permitindo que o leitor compreenda melhor o que foi lido. As pesquisas de Rosenthal e Ehri (2008) apoiam a teoria de que a visualização das palavras previamente à leitura do texto é benéfica para a compreensão textual.

### 3. Teoria da Aprendizagem Significativa

A teoria da aprendizagem significativa de Ausubel (1963) é baseada na premissa que a aprendizagem pode ser de dois tipos: *significativo* e *mecânico*. Aprendizagem significativa ocorre ao se assimilar novos conceitos e proposições no *corpo de conhecimento* (também chamado de “*estrutura cognitiva*”) que a pessoa já possui. A aprendizagem significativa é aquela em que as novas informações apresentadas ao indivíduo interagem com aquelas já consolidadas no seu corpo de conhecimento (Ausubel, Novak e Hanesian, 1978). Estas interações ocorrem entre a informação nova e aquela já consolidada, chamada por Ausubel de “*subsunçor*” ou “*ideia-âncora*” (Moreira, 2010). Este termo se refere a um conhecimento prévio que permite dar significado a uma informação nova. Independente de como o indivíduo entra em contato com a nova informação, seja por *descobrimto* ou *exposição*, o relacionamento dela e sua aprendizagem significativa está associado à presença de um subsunçor. Esses subsunçores formam uma estrutura cognitiva hierárquica, dinamicamente inter-relacionada.

Por outro lado, a aprendizagem mecânica (em inglês, *rote learning*) é semelhante à memorização de uma informação, sem que seja necessário dar um significado a ela. Este processo permite reproduzir em curto ou médio prazo a informação. Um exemplo comum de aprendizagem mecânica é a memorização de números telefônicos. Se o número for discado repetidas vezes, a pessoa irá memorizar tal informação; ao passar do tempo. Se a ação de discagem não for exercitada, haverá um esquecimento dessa informação, pois ela não está “*âncorada*” a nenhuma outra.

Para a aprendizagem ser significativa, é necessário que o indivíduo tenha condições de compreender a informação a ele apresentada e relacioná-la com informações já existentes em sua estrutura cognitiva. Para isso, duas condições devem ser preenchidas: o indivíduo deve possuir um conhecimento prévio do assunto de forma a poder relacionar o conceito novo com informação já conhecida (utilizando subsunçores) e a informação deve ser passada de forma clara e acessível. Ausubel (1968) considera que exemplos e relacionamentos explícitos do conceito são muito importantes nessa etapa, pois ajudam a pessoa a compreender o que está sendo ensinado e também lhe auxiliam a relacionar a ideia com conhecimentos prévios. Se a pessoa não

conseguir dar um significado próprio à informação, não será capaz de aprender significativamente.

Outro fator que influencia o processo de aprendizagem significativa é a vontade da pessoa de aprender aquele conceito ou ideia. Não é possível forçar um conhecimento para alunos; é necessário que eles tenham vontade de aprender e realizem uma ação interna de aprendizagem. Entre todos os fatores que afetam a aprendizagem significativa, Novak e Canas (2007) argumentam que a vontade de aprender é o fator mais difícil de ser influenciado pelo instrutor. As formas como o conteúdo é apresentado é fundamental para estimular o indivíduo a querer aprendê-lo, formalizando uma das condições necessárias para aprendizagem ser significativa. Técnicas de ensino que incentivam o estudante a relacionar o conteúdo em questão com os conhecimentos já estruturados por ele ou que o encorajam a pensar de forma crítica sobre alguma questão favorecem a aprendizagem significativa. Testes padronizados, com respostas curtas ou objetivas, favorecem a aprendizagem mecânica e podem inibir a vontade do aluno de aprender de forma significativa (Holden, 1992 apud Novak e Canas, 2007). Ainda de acordo com os autores, a aprendizagem mecânica contribui muito pouco para a estrutura de conhecimento do indivíduo e, portanto, não fomenta o pensamento criativo ou aptidões relacionadas à resolução de problemas.

De acordo com Ausubel e colaboradores (1987), é possível auxiliar os estudantes que estão aprendendo informações novas (e que não possuem subsunções para ‘ancorar’ tais informações) com organizadores prévios. Estes organizadores podem ser perguntas, situações, organizadores gráficos, entre outros inúmeros recursos disponíveis. Não há uma definição exata e única sobre o que são os organizadores prévios, mas eles são separados em dois tipos: *expositivo* e *comparativo*. O organizador *expositivo* é indicado quando o indivíduo não tem qualquer conhecimento sobre os conceitos a serem estudados e serve de ligação entre uma informação já conhecida e as informações novas que estão sendo apresentadas. O organizador *comparativo* pressupõe que a pessoa já possui na sua estrutura cognitiva conhecimentos similares e lhe ajuda a diferenciar o conhecimento novo daqueles já existentes.

Caso a aprendizagem significativa não ocorra, ainda é possível que a pessoa aprenda de forma mecânica uma informação. Porém, a transformação da aprendizagem mecânica em aprendizagem significativa não ocorre de forma automática. A pessoa que

incorporou mecanicamente novas informações em seu corpo de conhecimento, somente poderá aprender estes fatos de forma significativa se possuir algum subsunçor capaz de relacionar estas informações ao conhecimento já existente na sua estrutura cognitiva. A aprendizagem mecânica é feita de forma rápida pelo indivíduo, enquanto a significativa é progressiva e depende da captação e produção de significado (Gowin, 1981, Apud Moreira, 2010). O indivíduo deve internalizar um conceito e essa ação pode levar bastante tempo. Não basta a informação estar bem explicada para acelerar o processo, pois depende do conhecimento prévio da pessoa.

O conhecimento prévio do indivíduo é tão determinante para a aprendizagem significativa que foi considerada por Ausubel o fator mais importante para isso. De acordo com Ausubel (1963), o fator que mais influencia na aprendizagem é o conhecimento prévio do indivíduo. Ou seja, quanto maior for o corpo de conhecimento da pessoa, mais fácil será sua aprendizagem. A aprendizagem significativa ocorre justamente no encontro e relação dos conceitos novos e antigos, aumentando a rede de conhecimento e provendo maior estabilidade cognitiva para as informações já conhecidas pelo indivíduo.

A rede de conhecimento formada ao longo da vida de uma pessoa não é imutável. A teoria da aprendizagem significa não está apenas relacionada às novas informações e conceitos novos apresentados ao indivíduo. Aquelas informações já presentes no corpo de conhecimento podem sofrer alterações ao longo de sua existência. Uma das formas disso ocorrer é quando um novo conhecimento passa a ser subsunçor de um conhecimento prévio. Por exemplo, o conceito “*mapa*” remete inicialmente às conexões com “*mapa geográfico*”. Porém, ao longo da vida, esta ideia de “*mapa geográfico*” se torna uma especificação do conceito “*mapa*”, que pode ter muitos significados e definições. Este subsunçor “*mapa geográfico*”, que inicialmente foi utilizado para ancorar novas ideias, se torna um conceito ancorado pelo subsunçor “*mapa*”. As modificações em um subsunçor podem ser *derivativas* (quando a informação aprendida reforça aquela existente em um subsunçor) ou *correlativa* (quando a informação é uma extensão ou uma diferenciação da ideia presente no subsunçor).

É possível também que um subsunçor muito rico (com muitos significados claros e bem ancorado) seja pouco utilizado e perca seu significado para o corpo de

conhecimento. Quando uma informação não é utilizada por um espaço significativo de tempo, essa informação pode perder seu significado e a pessoa pode removê-la de sua base de conhecimento. A esse processo se dá o nome de “*assimilação obliteradora*” e demonstrar que, mesmo a aprendizagem ocorrendo de forma significativa, a informação pode ser ‘*esquecida*’ pelo indivíduo. Porém, é importante ressaltar que uma vez que a pessoa aprendeu significativamente um conceito, é mais fácil para ela retomar esse conhecimento, pois já há outras informações relacionadas ao conceito presentes em sua base de conhecimento. Se a aprendizagem do conceito for do tipo mecânica, é mais provável que o esquecimento do conceito seja total.

Dizer que o conhecimento prévio do indivíduo é a variável mais relevante para aprendizagem significativa não quer dizer que essa variável seja sempre um fator positivo. É possível que o conhecimento prévio funcione como um obstáculo para a aquisição de novos conhecimentos. Esses casos ocorrem quando o conhecimento aprendido pela pessoa está incorreto ou está demasiadamente ancorado a um contexto específico (um conjunto de informações) e a informação nova não está de acordo com o conhecimento previamente estabelecido. Este entrave na aprendizagem significativa se assemelha a um “obstáculo epistemológico” (Bachelard e Godinho, 1971 apud Moreira, 2010). A pessoa tem dificuldade em compreender uma informação nova por ela não estar de acordo com as demais informações de sua estrutura cognitiva.

Para considerar que a aprendizagem significativa acontece quando uma informação é compreendida e relacionada às demais informações já conhecidas do indivíduo é necessário identificar como ocorre o aprendizado dos primeiros conceitos, ou seja, os primeiros subsunçores. De acordo com Macnamara (1982), até os três anos de idade a criança reconhece conceitos através da repetição de padrões e observação do mundo ao seu redor. A experiência com objetos e eventos serve como estímulo para aquisição de conhecimento. Estes conceitos são aprendidos através do processo de *aprendizagem por descoberta* e formam a base do corpo de conhecimento do indivíduo. Após esta idade, a aprendizagem ocorre em grande parte pelo processo de “*recepção de conhecimento*”, que ocorre através de perguntas e interação com outros indivíduos. A diferença entre a aprendizagem por descoberta e a aprendizagem receptiva é a forma como a informação chega até o aprendiz. É importante ressaltar que nenhuma das duas é realizada de forma passiva; ou seja, é o aprendiz que necessita interagir com a fonte de informação e tenha vontade de aprender.

A partir das informações obtidas na infância, novos aprendizados são inseridos no corpo de conhecimento, utilizando as informações anteriores como subsunçores. Considerando o processo de construção do conhecimento descrito por Ausubel, torna-se fundamental que o conhecimento dos indivíduos seja levado em consideração no âmbito escolar e educacional. Uma das grandes dificuldades encontradas em sala de aula é como estruturar o curso para que sejam considerados os conhecimentos prévios dos estudantes e como relacionar as diferentes informações que se deseja que estejam presente no corpo de conhecimento deles ao final do curso. As estratégias de ensino devem levar em consideração tais fatores, tanto para que os estudantes possam relacionar as informações novas às já conhecidas quanto para fomentar a vontade dos alunos de aprenderem (um dos aspectos essenciais para a aprendizagem ser significativa).

Considerando o contexto escolar, muitos autores argumentam que as formas de ensino estão mais direcionadas para a aprendizagem mecânica que para a aprendizagem significativa (Brown, Collins, & Duguid, 1989 apud Novak 2002). Isso dificulta a tarefa de transferência do conhecimento para áreas diferentes. Para Novak (2002), esse é um dos principais problemas encontrados na atual forma de ensino, pois esses conhecimentos não são incorporados à estrutura de conhecimento do aluno, o que dificulta que sejam aplicados em diferentes cenários ou que sejam recordados em longo prazo.

Com o intuito de remediar essa situação, cresceu de forma substancial nas últimas décadas as pesquisas com uso de ferramentas para representação gráfica do conhecimento (Novak 2002). De acordo com Novak e Canas (2007), ferramentas que permitem organizar e relacionar conceitos de forma visual são propícias para auxiliar na aprendizagem significativa. Estas ferramentas permitem identificar conjuntos de ideias que são válidas ou que são erroneamente relacionadas, além de auxiliar a fomentar o relacionamento entre conceitos. As pesquisas de Marzano e colaboradores (2001) demonstram que organizadores gráficos auxiliam principalmente em tarefas que envolvam identificação de padrões, tais como informações de causa e efeito ou sequência de eventos. Já Novak e Canas (2007) e Moreira (2010) advogam que, através de organizadores gráficos, é possível produzir materiais que sejam “*potencialmente significativos*” para os estudantes. Não é possível inferir que um determinado material será completamente “*significativo*”, pois o significado dos materiais é dado pelo aluno.

Contudo, tais organizadores podem auxiliar os estudantes a ativar o conhecimento que já possuem e estabelecer relacionamentos entre este conhecimento e novas informações a eles apresentadas.

Dentre as ferramentas desenvolvidas para representação de conhecimento, *mapas conceituais* se destacam pela popularidade e extensão das pesquisas realizadas. Moreira (2010) expressa que mapas conceituais são ferramentas que podem ser utilizadas tanto como recursos instrucionais quanto para avaliação de estudantes. De acordo com o autor, através de um mapa conceitual o aluno demonstra como estão organizados os conceitos e os relacionamentos entre conceitos de um determinado tópico. Apesar de não ser uma representação precisa da estrutura cognitiva do aluno, os mapas conceituais fornecem os instrumentos necessários para a representação gráfica de ideias e relacionamentos entre elas. Porém, é importante ressaltar que não há um mapa conceitual *certo* ou *errado*; ele apenas reflete a representação de conhecimento da pessoa sobre determinado tópico e está associado à representação interna das informações que tenta expressar. Diversas pesquisas apontam que o uso de mapas conceituais durante o ensino de Física e Matemática podem melhorar o desempenho dos estudantes em testes padronizados (Novak, 2002).

A representação gráfica de conhecimento também tem sido utilizada para apoiar leitores a melhor compreenderem as informações contidas em textos. Chang e colegas (2002) consideram que a representação gráfica pode compor estratégias de leitura que diferem da leitura tradicional (linear). Informações extraídas do texto podem ser mais bem exploradas e interpretadas por estudantes na forma de grafos, nos quais as principais ideias e conceitos presentes no texto são representados como nodos, e os relacionamentos entre esses nodos são visualizados como as arestas (Chein & Mugnier, 2008). Nesta forma de representação da informação, as arestas não estão atreladas a um único tipo de informação, permitindo que o leitor interprete o relacionamento entre os conceitos de acordo com as informações presentes no texto. A interpretação das arestas e identificação dos termos presentes nos nodos exige que o leitor reflita sobre o texto, favorecendo o processo de aprendizagem significativa.

O uso contínuo de organizadores gráficos para compreensão textual tem demonstrado bons resultados, indicando que estudantes com baixa capacidade de leitura são aqueles que mais podem se beneficiar destes instrumentos (Manoli, 2012). Porém,

mesmo estudantes com boa capacidade de leitura podem se beneficiar do uso de tais ferramentas. Mayer e colegas (1996) apresentaram estudos que demonstram que um leitor capaz de ler com destreza textos científicos pode não conseguir transferir o conhecimento adquirido a outras áreas. Neste mesmo estudo, é demonstrado que a utilização de organizadores gráficos e ilustrações junto à leitura do texto melhorou o resultado dos alunos nos testes realizados e os ajudou na resolução de problemas relacionados às informações contidas no texto. Os autores finalizam apontando que esse processo auxiliou os estudantes a aprenderem de forma significativa os conteúdos apresentados por reduzir a carga cognitiva do processo de leitura, facilitando o processo de aprendizagem.

### **3.1 Considerações sobre o capítulo**

Este capítulo abordou o tema da aprendizagem significativa proposta por Ausubel (1963), trazendo a visão de Novak e Canas (2007) sobre as principais dificuldades encontradas em aplicar a teoria ao ensino escolar. É descrito como o conhecimento do indivíduo influencia na aprendizagem de novas informações e sobre como esse conhecimento é alterado ao longo do tempo. Métodos de ensino que exploram as características necessárias para que o aprendiz aprenda significativamente e ferramentas utilizadas para esse propósito também foram debatidas.

Ao final, foram apresentadas algumas ferramentas que se baseiam nesta teoria, como os organizadores gráficos e, mais especificamente, os mapas conceituais. Estes são ferramentas gráficas que possibilitam aos estudantes organizar e representar ideias por meio da explicitação de conceitos e seus relacionamentos, de maneira similar à forma como é proposta a utilização da ferramenta de mineração de textos apresentada nesta tese. Contudo, os mapas conceituais são normalmente construídos começando-se do zero, enquanto a proposta nesta tese é que os alunos partam de uma representação de conhecimento inicial criada pela ferramenta de mineração de texto. Ao apresentar ao aluno um organizador gráfico contendo um conjunto de termos de um texto que ele acabou de ler, criamos possibilidades de reflexão e reconstrução da representação das principais ideias do texto que fazem com que o aluno ative seu conhecimento prévio e busque no texto informações que possam ser ancoradas no que ele já conhece. Tal processo reflexivo tem suporte na teoria da aprendizagem significativa e propõe estratégias de leitura/compreensão textual que colocam o estudante numa postura ativa.

A pesquisa desenvolvida buscou demonstrar em que medida tais ferramentas tecnológicas e estratégias pedagógicas podem contribuir com o processo de leitura e compreensão textual.

## 4. Mineração de Texto

A mineração de textos é uma área de pesquisa em constante crescimento, principalmente devido ao aumento de informações que podem ser obtidas por meio de Internet. De livros à página pessoais, o conteúdo em forma de texto presente na web precisa ser classificado e catalogado, geralmente através de processos autônomos de análise e validação de dados. A forma como são realizados esses processos varia de acordo com a necessidade, mas geralmente está baseado em análises estatísticas e verificação de domínio, tendo análises semânticas e sintáticas como apoio.

Este capítulo apresenta uma revisão bibliográfica sobre mineração de textos, as principais abordagens para o assunto e suas aplicações.

### 4.1 Conceitos de mineração de textos

A definição de mineração de texto varia de acordo com o pesquisador que a propõe, mas é de consenso geral que a área está relacionada à mineração de dados, tendo como diferencial a possibilidade de análise de dados semiestruturados ou não estruturados (Feldman e Sanger, 2006; Tan, 1999). De acordo com Feldman e Sanger (2006), mineração de textos (*MT*) busca extrair informações consideradas úteis através da identificação e avaliação de padrões de dados não estruturados ou semiestruturados. Já Tan (1999) expande essa definição ao considerar que mineração de textos também deve ser capaz de extrair padrões *não triviais* dos dados apresentados e pode ser visto como uma extensão de área de pesquisa de mineração de dados, e semelhantes à área de pesquisa de *descoberta de conhecimento*. Gupta e Lehal (2009) define MT como sendo a descoberta de informações inéditas através de processamento de escrita. Para Bolasco e colegas (2005), mineração de textos é uma área de pesquisa que busca obter informações relevantes de textos que utilizam linguagem natural através de sua análise e estruturação (esta definição não leva em consideração métodos de análise estatísticos que não necessitam de linguagem natural para serem implementados).

Outras nomenclaturas para mineração de texto incluem "Análise Inteligente de textos", "Mineração de dados em textos" e "Descoberta de conhecimento" (Gupta e Lehal, 2009). É comum entre os autores de pesquisa nessa área considerar mineração de texto como uma área de pesquisa *interdisciplinar*, que utiliza pesquisas das áreas de

mineração de dados, aprendizagem de máquinas, estatísticas e processamento de linguagem natural (Tan, 1999; Gao et. al., 2005, Bolasco et. al., 2005).

Muitas das teorias e algoritmos de mineração de texto provêm da área de mineração de dados. Por isso, há uma grande similaridade nos processos realizados entre os dois campos de pesquisa. Ambos realizam pré-processamentos para organizar os dados, buscam encontrar padrões nesses dados e apresentam formas semelhantes de visualização das informações encontradas. Além disso, algoritmos utilizados em MT geralmente possuem heurísticas desenvolvidas inicialmente para mineração de dados quando se trata de encontrar conhecimento na sua base de dados. Tan (1999) considera que mineração de textos é um processo mais complexo que mineração de dados por envolver dados não estruturados e ambíguos.

O processo de MT primeiro deve transformar um documento (textual) em uma forma estruturada de dados. Essa transformação utiliza alguma característica ou marcação do documento para estruturá-lo de forma que o processo de extração de conhecimento possa ser realizado. Um grafo conceitual ou uma representação do tipo base de dados relacional é geralmente utilizado (Tan, 1999). Após esse processo de transformação de dados, um segundo algoritmo encontra padrões nos dados já estruturados. Um exemplo desse processo é visto na Figura 1.

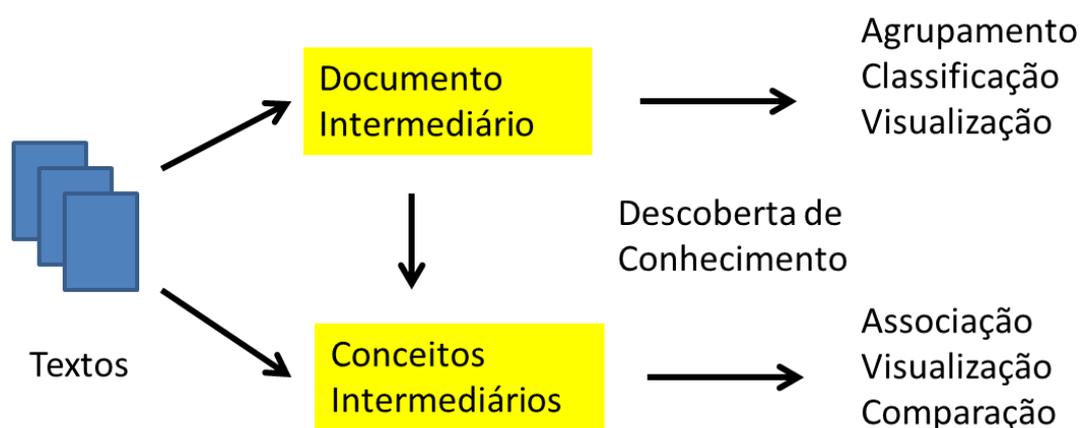


Figura 1 Processo de mineração de textos. Figura adaptada de Tan (1999)

O pré-processamento realizado pelos algoritmos de mineração de texto é responsável por transformar documentos não estruturados (ou semiestruturados) em dados estruturados. Este aspecto difere bastante MT de mineração e dados, uma vez que no segundo caso os dados necessariamente já estão estruturados. Os documentos

processados por um minerador de textos são definidos de forma informal por Feldman e Sanger (2006) como “*unidades de dados textuais*”. Exemplos de documentos são relatórios, memorandos, e-mails, artigos, noticiais, etc. Um documento também pode ser inserido dentro de um contexto e analisado com base no contexto.

A análise e estruturação de dados de um documento podem ser realizadas utilizando marcados ou características consideradas como estruturas do texto. Os quatro tipos de características mais utilizadas para demarcar dados no texto são: caracteres, palavras, termos e conceitos (Feldman e Sanger, 2006).

- **Caracteres:** são os componentes individuais das palavras, tais como letras, números e símbolos. Eles são os menores componentes do texto e formam os demais tipos de marcadores (palavras, termos e conceitos). A extração de caracteres de um texto não é um processo muito utilizado por não apresentar informações que sejam relevantes à compreensão do texto e dos dados presentes nele. Porém, algoritmos que extraíam bigramas ou trigramas são comuns e possuem algumas finalidades, como identificar o idioma em que um texto está escrito.

- **Palavras:** a extração de palavras de um documento sem nenhum tipo de pós ou pré-processamento pode ser considerada como a forma mais básica de extração de conhecimento do texto. Ao se utilizar *palavras* como marcadores, considera-se que cada palavra contida no texto corresponda a um “*token*”. Frases, expressões e palavras com hífen não são reconhecidos durante a mineração. É possível também que documentos contenham milhares de palavras diferentes, cada uma sendo considerada um *token* diferente. Nesses casos, a extração de palavras poderia gerar uma base de dados demasiadamente densa para ser avaliada pelo usuário. Por isso, mesmo esta sendo considerada uma forma básica de MT, a maioria dos mineradores que realizam a extração de informações baseada em palavras utiliza alguma heurística para impedir que todas as palavras encontradas sejam listadas como resultado da mineração. Uma otimização simples para reduzir o espaço de resposta do minerador é desconsiderar palavras contidas no conjunto de “*stop words*” (grupo de palavras comuns que não adicionam informação semântica, basicamente formada por artigo e proposições) e numerais.

- **Termos:** podem ser constituídos de uma única palavra ou de um conjunto de palavras. A mineração de texto por termos utiliza unicamente o corpo do documento para identificar os termos extraídos e julgar como eles devem ser formados. Assim, o

conjunto de termos extraído de um documento deve ser composto exclusivamente por termos que estejam presentes no documento. Nenhum tipo de auxílio externo (como diferentes documentos ou ontologias) é utilizado nesse tipo de mineração. Este tipo de processo geralmente apresenta um resultado com mais informações semânticas que apenas a mineração de textos utilizando palavras. Além disso, expressões, palavras separadas por hífen e frases também podem ser encontradas como resultado da mineração. Algoritmos de pós-processamento e heurísticas podem ser utilizados para reduzir o número de termos resultantes da mineração e encontrar aqueles que são mais relevantes para a compreensão do texto. Uma forma comum de reduzir a quantidade de termos apresentados após a mineração é utilizar a frequência com que cada termo ocorre no documento para determinar os mais recorrentes, sendo eles possivelmente os mais relevantes para a identificação do texto.

- **Conceitos:** são elementos gerados para um documento através de análise manual, estatística, ou com base em regras, ou abordagem híbrida (Feldman e Sanger, 2006). Apesar de esses elementos poderem ser gerados de forma manual, é cada vez mais comum que sejam gerados por pré-processamento que identifica certos termos e expressões do texto e comparam-nos a um conjunto de dados específicos. Ao se classificar a mineração como sendo capaz de extrair conceitos de um texto, é comum que se use fontes externas ao documento (como base de dados, ontologias e outros textos) para encontrar os conceitos. Por exemplo, um conjunto de documentos que não inclua a palavra "*dados*" em nenhum texto pode ainda assim ter esse conceito sugerido por um minerador. Isso ocorrerá na hipótese desse conjunto de textos ser identificado como parte de um grupo que contém o conceito "*dados*" como uma de suas palavras-chave.

Dentre as diversas abordagens e algoritmos utilizados pela MT, os mais comuns incluem heurísticas voltadas à análise de distribuição e frequência dos marcadores do documento para extrair as informações consideradas relevantes. A relação de cada informação com o documento (e com as demais informações) por vezes é obtida considerando o escopo do documento e a sua relação com outros documentos. Por exemplo, se um documento acerca de um remédio estiver relacionado com outro documento que debate uma determinada doença, a relação entre as informações extraídas dos documentos pode indicar que o remédio apresentado no primeiro documento está relacionado com a doença discutida no segundo documento. Este tipo

de relacionamento pode ser realizado dentro de um mesmo documento ou como estratégia de análise de um conjunto de documentos.

Este tipo de mineração de textos e de *extração de informação* permite encontrar informações de forma implícita. Esse processamento está além da simples busca por termos ou conceitos que revelem o conteúdo de um documento. O resultado deste tipo de extração de informação pode auxiliar na compreensão dos dados contidos nos documentos e permitir uma análise de como eles se relacionam. Quando considerado um conjunto de documentos, a relação entre eles e entre seus conceitos cria uma representação mais rica e abrangente de seu conteúdo. Assim, o processo permite compreender determinadas características dos documentos e pode ser visto como uma forma de entender o texto de forma completa (Feldman e Sanger, 2006). O processo também busca encontrar relacionamento entre os termos, representando de forma mais completa o conteúdo do documento. O número de relacionamentos e termos do texto é muito maior que o número de palavras-chave extraídas (uma vez que é necessário considerar todos os relacionamentos entre termos), sendo necessário realizar essa extração de informação de forma automatizada e utilizando heurísticas para remover as relações consideradas menos relevantes.

## **4.2 Aplicações da mineração de textos**

As informações encontradas na internet estão na sua grande maioria presentes no formato de texto e dados semiestruturados, enfatizando a importância das pesquisas na área de mineração de texto. De acordo com os pesquisadores Tan (1999) e Gupta e Lehal (2009), cerca de 80% dos dados presentes na internet estão na forma de texto. Os dados na forma de texto não podem ser minerados com algoritmos desenvolvidos para mineração de dados.

MT vem sendo aplicada a diferentes áreas de pesquisa e em diferentes situações. Os algoritmos atuais de mineração permitem a extração de informações muito rapidamente, sendo apropriados para análise de grandes quantidades de textos e documentos. Isso contribui para que empresas e companhias utilizem MT tanto para gerência de trabalho interno quanto para análise de dados. Gupta e Lehal (2009) aponta que MT é utilizada como auxílio à área de recursos humanos das empresas, através de análise de opiniões, resultados, satisfação dos empregados e gerência de currículos.

Também é possível analisar e classificar o contato dos clientes com empresas, através de algoritmos que encontram as informações mais relevantes desses contatos e repassa essas informações para o setor apropriado (Bolasco et. al., 2005).

Outra importante área de atuação de MT é a área comercial. Nestas áreas, a MT está diretamente associada à análise de mercado e descoberta de conhecimento de mercado. Através da avaliação das ações de outras companhias, do público e das mídias, empresas tentam prospectar novos clientes, investimentos e ações de marketing (Gupta e Lehal, 2009).

Outros autores propõem a utilização de MT para análise de patentes (Tseng et. al., 2007), comércio eletrônico (Pang e Lee, 2008), para prever o valor de ações na bolsa de valores (Geva e Zahavi, 2010) e analisar a condição de pacientes na área da medicina (Damasceno et. al., 2004). De acordo com Bolasco et. al. (2005), as principais áreas de pesquisa em mineração de textos estão relacionadas com:

- Mídia
- Telecomunicações
- Tecnologia de informação
- Recuperação de informações na Internet
- Bancos e seguradoras
- Mercado financeiro
- Política
- Grupos e pesquisas farmacêuticas
- Educação e cursos a distância

No âmbito da área educacional, a mineração de textos vem se destacando principalmente nos últimos anos. Com o crescimento dos cursos a distância e dos *Cursos Online Aberto e Massivo (MOOC)* surgiu a necessidade de processar e analisar uma quantidade sem precedentes de textos e trabalhos. Muitos autores já utilizam mineração de textos para auxiliar essa tarefa, seja na forma de análise de postagem em fóruns (Azevedo et. al., 2014), síntese de textos para auxiliar a escrita colaborativa (Macedo et. al., 2009) ou mesmo incorporada a agentes pedagógicos (Pinho et. al., 2013).

Outra utilização para mineração de textos diz respeito à extração de palavras-chave de textos (*keywords*). Há sites que utilizam MT para extrair de forma automática as palavras-chave mais relevantes de cada artigo publicado para indicar tais conteúdos aos seus usuários, com base em escolhas pré-definidas realizadas por eles. Isso permite a indicação de um conteúdo mais personalizado, que seja de interesse do usuário. Devido ao alto número de artigos e documentos publicados a cada dia, essa tarefa de identificação de palavras-chave não poderia ser realizada de forma manual. Buscadores online também utilizam a extração de palavras-chave tanto para encontrar resultados coerentes com a busca quanto para realizar a categorização dos sites e conteúdos.

A categorização de documentos classifica textos com os mesmos conceitos ou assuntos em grupos. Quando se categoriza um documento, o processo de MT o trata como sendo um conjunto de palavras, sem se importar com a semântica ou relação entre elas (Gupta e Lehal, 2009). Para identificar de forma mais precisa a que grupo cada texto pertence, algoritmos de categorização costumam utilizar dicionário de sinônimos para situar cada termo encontrado dentro de um grupo de termos semelhantes. Outra característica desses algoritmos é que há uma ordenação entre os documentos categorizados. Essa ordenação pode ser realizada por diferentes equações, fornecendo um peso para cada aspecto do documento minerador e sendo utilizada ao se buscar documentos com características específicas.

Mineração de textos também é comumente utilizada para sumarização de textos. A construção de sumários para textos pode auxiliar na compreensão de documentos demasiadamente grandes ou verificar se um determinado documento contém o conteúdo que se está procurando. A sumarização de textos pode prover informações acerca de locais, tempo, pessoas e outros dados recorrentes no texto, mas é muito difícil envolver semântica nesse processo. Não há muitas heurísticas que consigam (com um razoável grau de confiança) identificar a relação semântica entre os conceitos extraídos da sumarização. Geralmente, para que esse processo tenha melhores resultados, os algoritmos de mineração de textos podem ser auxiliados por informações, ontologias ou dados que permitam verificar se há conexão entre determinadas palavras-chave. Algoritmos de sumarização costumam empregam técnicas estatísticas para extrair as informações consideradas mais relevantes, através da análise de recorrência dessas informações no texto (Gupta e Lehal, 2009).

Após a extração das informações do texto, é necessário apresentá-las ao usuário. A representação da informação exige que haja algum tipo de estrutura de dados e que essa estrutura possa ser facilmente compreendida. Para isso, muitos autores defendem a representação por meio de grafos, pois estes permitem organizar as informações e estabelecer relações entre elas (Chein e Mugnier, 2008). Uma vez que grafos são representações abstratas, a relação entre as informações e a forma como elas são apresentadas podem variar para melhor se adaptar às necessidades dos usuários. Outra característica importante do uso de grafos é que as relações entre os dados apresentados podem ajudar o usuário a melhor entender o significado dos mesmos.

Grafos são construções matemáticas importantes e eficazes para modelar estruturas de informação, permitindo que haja uma conexão entre os dados. A sua utilidade é tão abrangente que há uma série de algoritmos específicos para operações em grafos e é comum mapear problemas do dia-a-dia para problemas em grafos. Alguns exemplos de problemas mapeados para grafos são os de ordenamento, de compressão de dados, de uso de mapas, de rotas, de alocação de recursos, entre outros (Cormen et. al., 2007). Além de problemas comuns do dia-a-dia, questões tecnológicas comumente tratadas com uso de vetores também estão se apropriando de algoritmos de grafos para aumentar sua eficiência e capacidade. Reconhecimento facial (Wiskott et. al., 1997 apud Schenker, 2003), de digital (Shen et. al., 2001 apud Schenker, 2003) e grande parte da comunicação web (desde algoritmos de busca até navegação por websites) utilizam princípios de estruturas de grafos para aumentar sua eficiência (Dickinson et. al., 2002 apud Schenker, 2003).

Para realizar o mapeamento da informação obtida no texto em um grafo, é necessário estruturar a informação extraída. Um dos métodos mais tradicionais para isso é a associação de cada dado encontrado a um valor. Se for utilizado um valor binário, essa representação indica se há ou não um determinado dado em um documento. Também é possível utilizar valores inteiros, representando o número de vezes que cada dado aparece em um texto. Se forem utilizados números reais, esses valores representam o peso (normalizado) que cada um tem em um texto.

A representação através de números agrupados em um vetor, onde cada posição corresponde a um dado é chamada de “*Modelo Vetor-Espacial de Recuperação de Informação*” (Schenker, 2003). Este modelo possui característica matemáticas bastante

peculiares, sendo constantemente utilizado para aprendizagem de máquinas. Esta abordagem é amplamente utilizada para agrupamentos e coleção de documentos, uma vez que é possível estabelecer vetores de informação para mais de um documento e utilizá-los para trabalhar com informações existentes em vários textos.

### 4.3 Minerador de textos Sobek

A ferramenta utilizada como parte deste trabalho tem o nome de **Sobek**. Trata-se de um aplicativo de mineração de textos implementado utilizando linguagem de programação Java e desenvolvido como parte do trabalho de Lorenzatti (2007) e Macedo (2010). A ferramenta foi inicialmente idealizada para auxiliar professores a enfrentar problemas característicos do ensino a distância (Macedo et. al., 2009). De acordo com a pesquisadora:

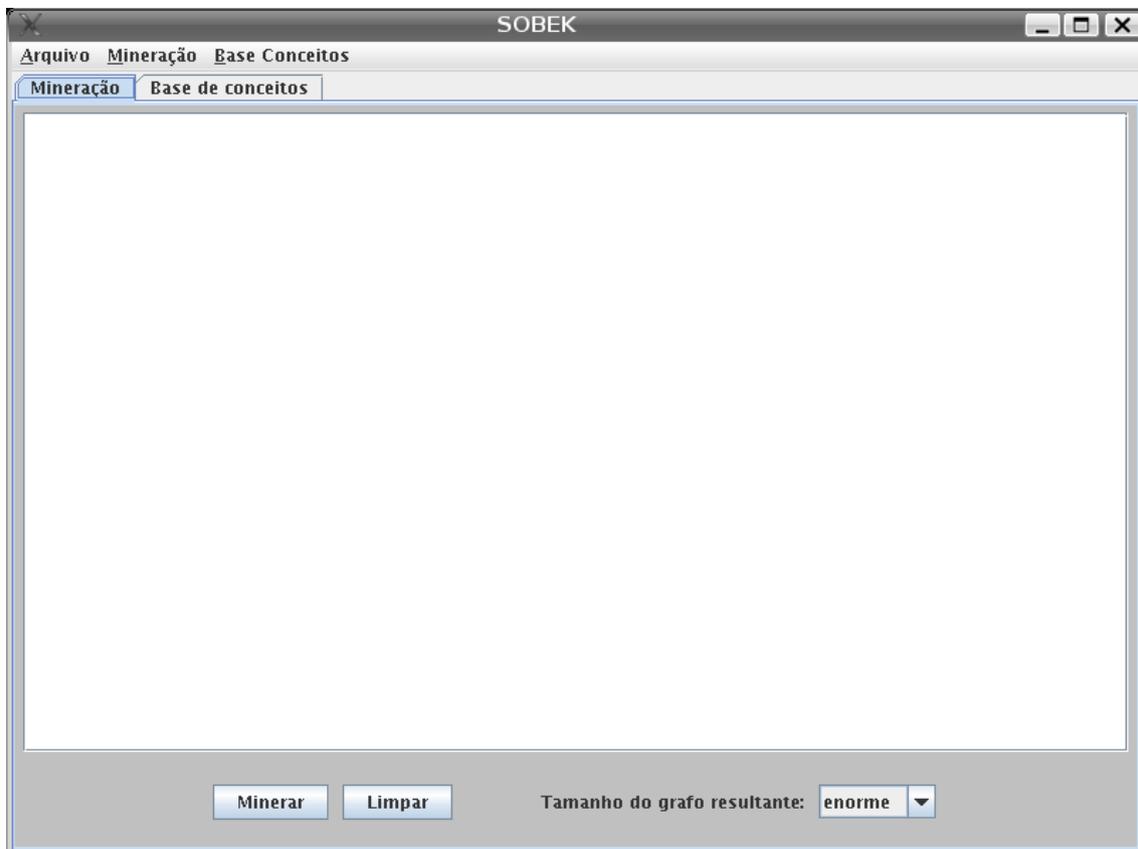
*“... o estudo identificou a necessidade de uma ferramenta que viabilizasse, em menor tempo, o acompanhamento e gerenciamento, pelo professor, do alto volume de dados gerados nas produções textuais coletivas a distancia.” (página 94)*

Macedo, 2010

Frente a uma sempre crescente quantidade de material escrito por alunos e disponíveis para pesquisa, além de um aumento expressivo no ensino a distância (*EaD*) nos últimos anos, percebeu-se que havia a necessidade de auxiliar professores que ministravam cursos EaD a classificarem e analisarem textos de maneira mais rápida e automática. A forma que se encontrou para fazer isso foi o desenvolvimento de um minerador de textos, que resultou no aplicativo Sobek. Ao fornecer a esses professores uma ferramenta para a análise de mensagens e contribuições dos estudantes, o Sobek torna-se uma ferramenta de apoio ao trabalho do docente, dando aos educadores mais tempo para se concentrarem em problemas específicos de suas áreas.

A ferramenta Sobek extrai informações de um texto e as apresenta como um grafo, no qual os principais termos extraídos do texto estão dispostos como vértices (nodos) e a ligação entre os termos como arestas. Isso permite representar o texto de forma gráfica e auxilia o usuário a visualizar os principais termos do texto e suas

relações. Diferente de outros mineradores de textos tais como o RapidMiner<sup>1</sup> e o IBM ManyEyes<sup>2</sup>, ele possui como principal característica uma interface simples, pensada para permitir o seu uso sem a necessidade de treinamento prévio ou de conhecimentos de informática.



**Figura 2 Interface de uso do Sobek. Retirado de Lorenzatti (2007)**

Ao mesmo tempo em que sua interface é simples e fácil de utilizar, surgiu a necessidade de adicionar à ferramenta funcionalidades que permitissem um maior controle por parte do usuário sobre as configurações e características da extração de termos. Com base nessas necessidades, foram implementadas configurações que permitem ao usuário um controle maior sobre as características e estratégias de mineração de texto utilizada pelo Sobek. Dentre as possíveis configurações do Sobek, destacam-se:

- *Seleção de frequência mínima*: Estabelece qual a frequência mínima que um termo deve ter para ser apresentado como termo relevantes do texto.

---

<sup>1</sup> <https://rapidminer.com/>

<sup>2</sup> <http://www-01.ibm.com/software/analytics/many-eyes/>

- *Número médio de termos*: Define um número mínimo de termos que devem estar presentes na relação final do Sobek.
- *Lista de stop words*: Permite adicionar, remover ou substituir a lista de *stop words* (palavras que não devem ser consideradas durante a mineração de texto) utilizada pelo minerador.
- *Idioma*: O Sobek possui suporte para os idiomas Português e Inglês. Está em desenvolvimento o suporte para idioma espanhol.

A interface inicialmente planejada para a ferramenta pode ser visualizada na Figura 3. A utilização do Sobek baseia-se na cópia de um texto para a área da ferramenta destinada à edição de texto. A facilidade de pressionar apenas um botão para extrair os termos considerados relevantes é uma das características importantes da ferramenta. Além de extrair termos de textos copiados na caixa de edição de texto, também é possível abrir um texto salvo no computador ou mesmo uma mineração realizada previamente. A ferramenta possui suporte para trabalhar com arquivos no formato “.doc” e “.pdf”.

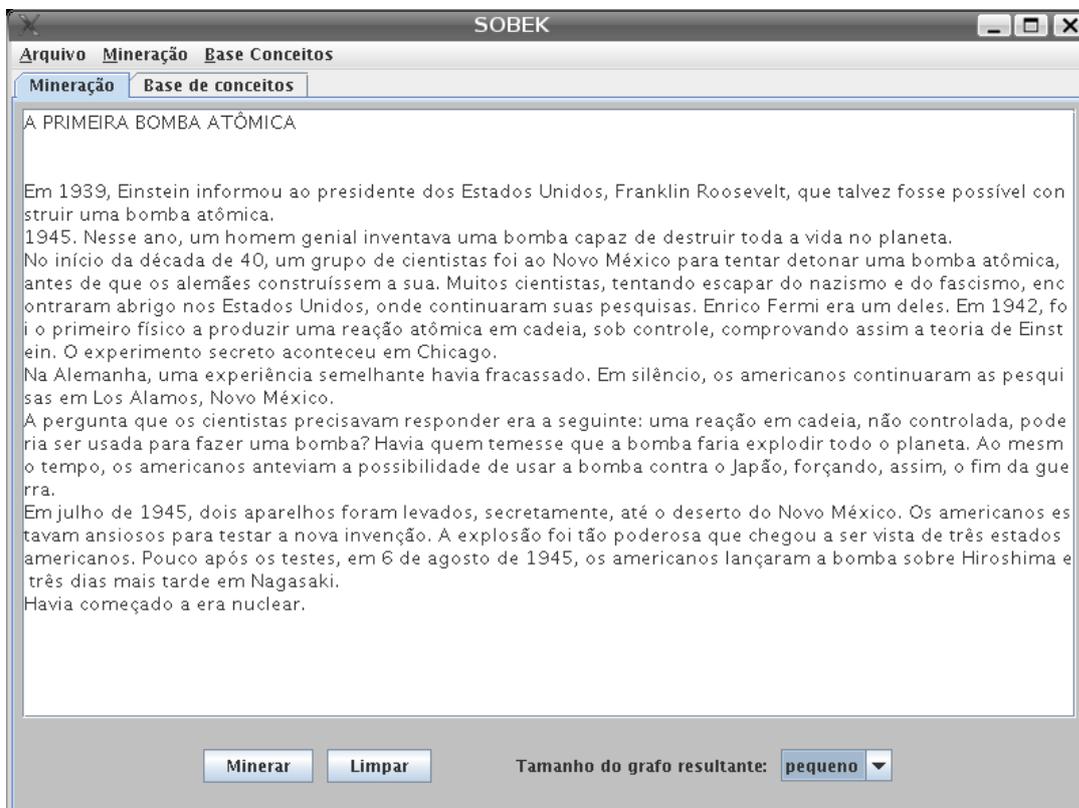


Figura 3. Interface do Sobek com texto. Retirado de Lorenzatti (2007)

O Sobek também pode analisar um conjunto de textos, mesmo se eles estiverem em diferentes formatos. Um dos usos para esse recurso é a possibilidade de se analisar um texto em particular para procurar termos encontrados em outros documentos. Por exemplo, um professor pode pedir a seus alunos que escrevam um texto com base em determinado artigo. Através do Sobek, a bibliografia dada aos estudantes pode ser utilizada para criar uma base de termos. Esta base de termos pode então ser utilizada para verificar se os estudantes estão discutindo questões contidas na bibliografia sugerida (Pinho et. al., 2013).

Uma vez que o Sobek tenha completado a extração de termos é necessário apresentar os dados extraídos ao usuário. A ferramenta busca apresentar os resultados de maneira clara e objetiva, tornando possível identificar rapidamente os termos mais importantes do texto e demais informações relevantes. O formato de representação gráfica escolhido foi o formato de grafo. A escolha pelo formato de grafo se deve pela fácil e simples visualização, possibilidade de conexão entre termos e a capacidade de adicionar informações (tais como número de ocorrências de um termo) sem a necessidade de outras telas ou visualizações (Schenker, 2003). A representação das informações extraídas do Sobek também busca assemelhar-se a de um mapa conceitual. Ela baseia-se na concepção de aprendizagem significativa proposta por Ausubel (1963) e procura auxiliar o usuário a melhor compreender a relação entre os termos e associar as informações obtidas do texto àquelas já conhecidas por ele.

Durante desenvolvimento da ferramenta, foi proposto que a rede de termos extraída pelo Sobek pudesse ser alterada previamente à visualização em grafo dos termos (Figura 4). Neste cenário, o usuário teria acesso aos termos extraídos e suas relações, podendo alterar ele conforme fosse necessário. Após a base de termos ter sido alterada, o grafo era gerado pelo usuário (Figura 5).

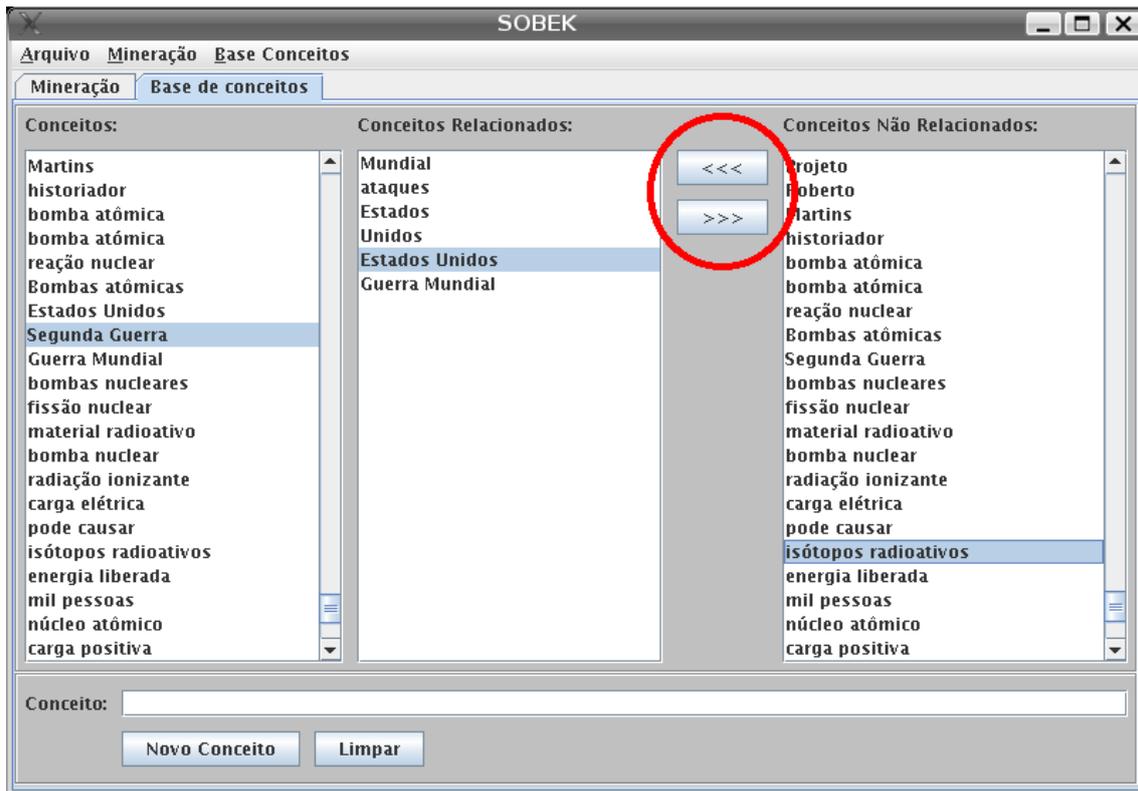


Figura 4 Representação da base de termos extraída pelo Sobek, com os termos e seus relacionamentos. Retirado de Lorenzatti (2007)

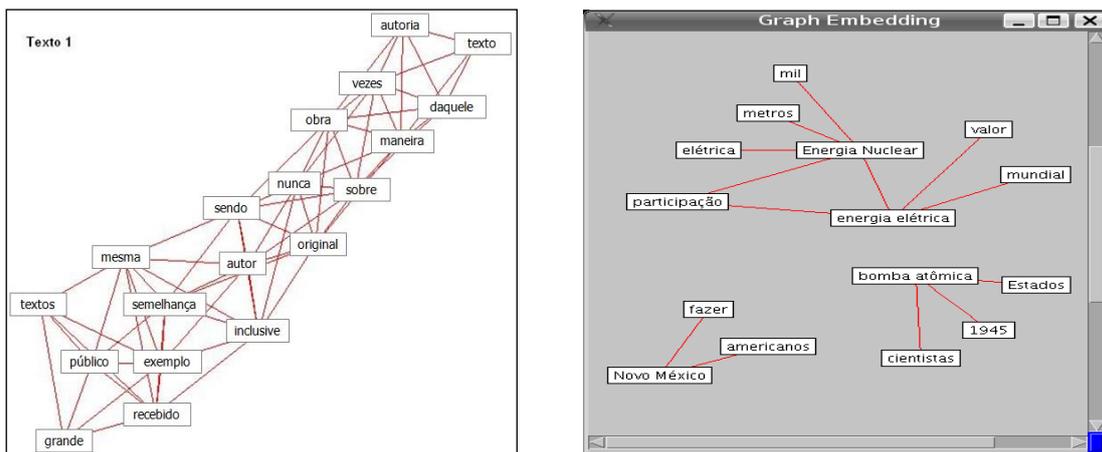


Figura 5 Representações de grafos produzidos pelo Sobek. À esquerda, imagem retirada de Macedo (2010); à direita, imagem retirada de Lorenzatti (2007).

O grafo resultante da mineração de texto apresenta os termos considerados mais relevantes de acordo com a frequência de ocorrência deles. Também são apresentadas

conexões entre os termos. Estas conexões buscam representar relações existentes no texto e podem indicar efeito de causa e consequência, relações temporais ou mesmo termos relacionados pelo seu significado.

Apesar de o Sobek ter uma interface gráfica simples de utilizar, surgiu a necessidade de aplicar o algoritmo do Sobek também em projetos que não contêm uma interface direta com o usuário. Nestes projetos, o Sobek é utilizado para buscar informações contidas no texto. Azevedo et. al. (2014) utilizou os dados extraídos pelo Sobek para elaborar um algoritmo capaz de encontrar similaridade entre postagens em fóruns e fornecer uma análise de similaridade entre eles. O trabalho do pesquisador utilizou tanto a lista de termos considerados relevante extraído pelo Sobek quanto o relacionamento entre esses termos. Já Damasceno et. al. (2014) propôs que os dados fornecidos pela mineração do Sobek pudessem ser utilizados para mapear problemas às soluções em um sistema de tele saúde. Acosta et. al. (2014) utilizou apenas os termos considerados mais relevantes pelo Sobek para buscar recomendação de conteúdo baseado nos textos escritos por estudantes.

Estas propostas diferem do estudo realizado nesta tese na medida em que ela tem como foco o uso da ferramenta Sobek para apoio à compreensão textual, tema não tratado anteriormente.

#### **4.4 Considerações sobre o capítulo**

A mineração de textos nos permite encontrar informações contidas em textos de forma automática. Estas informações são utilizadas em diversas áreas de pesquisa e para inúmeras aplicações. Na área pedagógica, é possível encontrar aplicações voltadas tanto para o auxílio ao professor quanto ao aluno.

Considerando os dados que se deseja encontrar durante a mineração de textos, Feldman e Sanger (2006) separam eles em 4 tipos: caracteres, palavras, termos e conceitos. Para cada tipo de dado há diversos algoritmos que podem ser descritos, desde analisadores sintáticos até analisadores semânticos. Os mineradores de textos também podem ser utilizados para encontrar informações dentro de um documento ou um conjunto de documentos, realizando relacionamentos entre eles.

Neste capítulo foram descritas as teorias e conceitos relacionados à mineração de texto, bem como demonstrado suas aplicações e limitações. Também foi apresentado o minerador de textos Sobek, desenvolvido para auxiliar professores nas tarefas relacionadas ao ensino à distância. O Sobek permite extrair termos e relacionamentos de textos e demonstrá-los na forma de um grafo. Através das técnicas de extração de termos e representação de informação, o Sobek torna-se uma ferramenta de mineração de textos capaz de auxiliar em diversas atividades pedagógicas. Nesta tese, está delineada uma pesquisa que visa utilizar o minerador Sobek para auxiliar estudantes em atividades de leitura e compreensão textual.

## 5. Metodologia

Este capítulo apresenta a metodologia de pesquisa utilizada nesta tese. Primeiramente, a ferramenta a ser utilizada durante a pesquisa é descrita. Nesta descrição, suas funcionalidades e principais características são enfatizadas, elementos que foram desenvolvidos com o intuito de auxiliar a leitura e a compreensão textual. Logo após, são apresentados estudos preliminares que buscam demonstrar a validade da ferramenta proposta para a realização das atividades de letramento, as quais permitiram avaliar pontos fortes e limitações do Sobek. Ao final, os delineamentos gerais dos métodos de experimentação da pesquisa são descritos, bem como as justificativas para as escolhas metodológicas.

Conforme já apresentado na introdução desta tese, o objetivo geral da pesquisa foi definido como: **Investigar como a ferramenta de mineração de textos Sobek pode apoiar o processo de leitura e compreensão textual.**

Decorrentes deste objetivo foram estabelecidos objetivos específicos:

- Aprimorar e avaliar a ferramenta Sobek no que diz respeito à sua capacidade de extração de termos e como elemento de apoio à leitura e compreensão textual;
- Elaborar práticas pedagógicas que podem ser utilizadas associadas ao Sobek a fim de potencializar sua utilização para apoio aos processos de leitura e compreensão textual.

A próxima seção descreve modificações e aprimoramentos realizados na ferramenta Sobek com o propósito de adequá-la a esta pesquisa.

### 5.1 Aprimoramento da ferramenta Sobek

Ao longo do desenvolvimento desta pesquisa, diversas funcionalidades e características do minerador Sobek foram alteradas para facilitar a compreensão dos resultados da mineração de textos, melhorar o algoritmo de extração de termos e relacionamentos, e facilitar a interação do usuário com a ferramenta. A seguir são descritas as alterações realizadas no minerador, seu algoritmo e estudos preliminares que buscam verificar a capacidade da ferramenta em encontrar termos considerados relevantes em um texto.

### 5.1.1 Alteração da ferramenta de mineração de textos Sobek

Durante a pesquisa, muitas alterações gráficas e funcionais foram desenvolvidas junto ao minerador de texto. As alterações gráficas contemplam modificações desde a tela inicial de interface com o usuário (Figura 6) até a representação do grafo dos termos e relacionamentos extraídos.

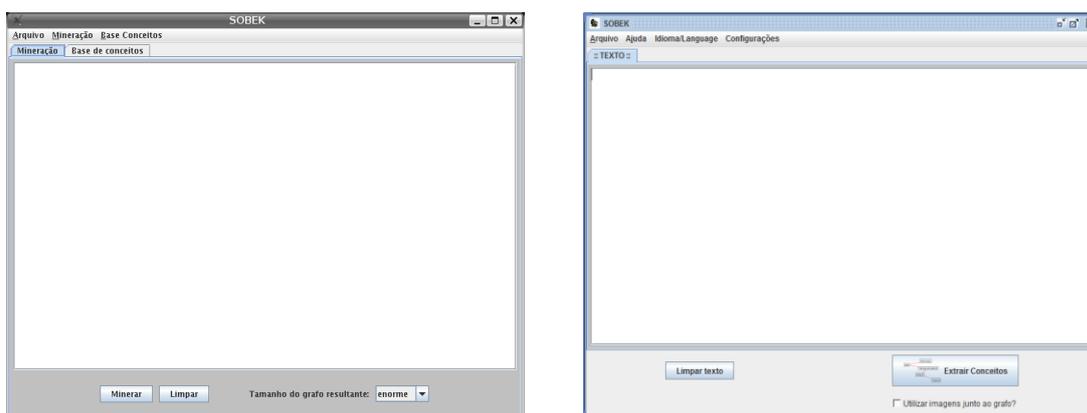


Figura 6 Interface original do minerador Sobek a esquerda (Lorenzatti, 2007) e interface remodelagem apresentada a direita.

À esquerda, observa-se a apresentação original da interface do minerador Sobek (Lorenzatti, 2007). À direita, a interface remodelada, produto deste trabalho, buscou facilitar a operação da ferramenta e disponibilizar funcionalidades adicionais, como a possibilidade de busca de imagens online que pudessem representar os termos extraídos do texto.

As alterações realizadas inicialmente modificaram a tela inicial do Sobek, buscando tornar o seu uso ainda mais simples e rápido. A etapa de representação e edição dos termos e relacionamentos extraídos pela ferramenta foi removida do Sobek, permitindo assim que o usuário visualizasse o grafo representativo do texto de forma imediata. Isso acarretou na necessidade de se alterar a representação do grafo, dada a necessidade do usuário poder alterar os termos e relacionamento diretamente nele. Outro motivo para a remoção da tela intermediária entre a extração de termos e sua visualização é que o usuário poderia não querer ver os termos ou não saber o que alterar na base de termos, estando apenas interessado em visualizar a representação do texto.

Isso facilitou o acesso da ferramenta a usuários com pouca experiência com tecnologia ou pouca familiaridade com o minerador.

Nesta nova interface do Sobek, o grafo de visualização de termos se tornou a peça principal da ferramenta, uma vez que todas as alterações das informações extraídas são realizadas diretamente no grafo. Diversas interfaces gráficas foram utilizadas e modificadas ao longo da pesquisa, sempre buscando melhorar a visualização e representação das informações. As primeiras alterações efetuadas na visualização dos termos no grafo podem ser encontradas na Figura 7, onde é possível observar que certos termos são destacados. Após diversas alterações buscando maior clareza na representação, a versão final do grafo apresentado pelo Sobek após uma mineração é apresentada na Figura 8. Dentre as características representadas nesse grafo, é importante mencionar que o tamanho, cor e formato de cada nodo estão associados ao número de ocorrências do termo no texto. Quanto mais recorrente o termo, maior o nodo que o engloba e mais escura é a sua fonte. Dessa forma, é possível rapidamente identificar em um grafo os termos mais recorrentes do texto. Do ponto de vista educacional, estas características deram à ferramenta a capacidade de auxiliar o leitor na identificação de termos e conceitos importantes em um texto, o que está diretamente relacionado à habilidade de compreensão textual.

A utilização do minerador de texto em diferentes projetos piloto também permitiu observar que a API escolhida (a *Interactive Graph Drawing*<sup>3</sup>) inicialmente para apresentar o gráfico ao usuário possuía um grande número de limitações, tais como baixo número de formas geométricas para visualização, dificuldade na seleção de nodos, impossibilidade de edição de termos, entre outros. Com base nessas observações, optou-se por alterar a API para outra mais recente, chamada de *Prefuse*<sup>4</sup>. Esta nova API de representação de grafos possui maior possibilidade de configuração e de interação com o grafo desenhado. Após esta alteração, foi possível apresentar ao usuário uma maior quantidade de informações referentes a cada termo e melhorar a sua representação no grafo.

---

<sup>3</sup> <http://www.cs.rpi.edu/research/groups/pb/graphdraw/>

<sup>4</sup> <http://prefuse.org/>

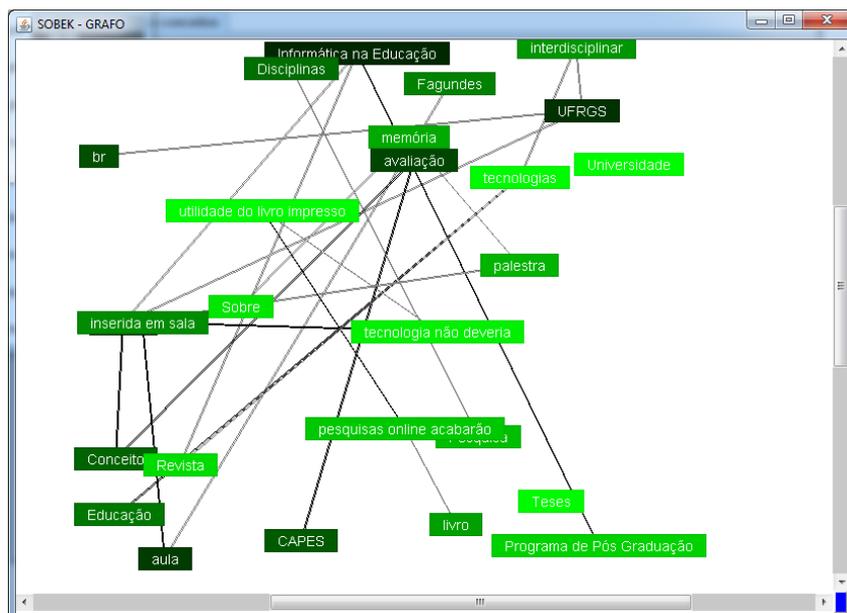


Figura 7 Visualização do grafo utilizando *Interactive Graph Drawing API* após alterações iniciais (imagem gerada pelo autor).

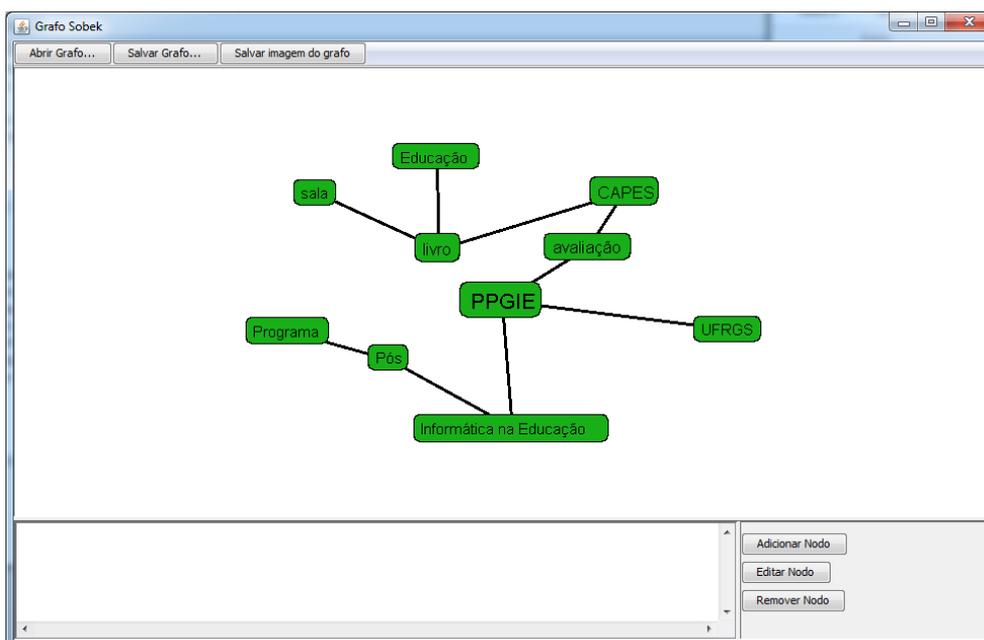


Figura 8 Visualização do grafo resultante da operação de extração de termos do Sobek. É utilizado a *Prefuse API* (imagem gerada pelo autor).

A nova representação dos termos extraídos pelo Sobek permite ao usuário interagir com cada nodo e com o grafo como um todo. Para permitir que o grafo seja personalizado pelo usuário, foram adicionadas funcionalidades de adição, remoção e edição de nodos. Isso serve como substituto da *base de conceitos* presente na antiga

versão do minerador. Além disso, permite ao usuário interagir de forma mais completa com o grafo. Esta interação normalmente ocorre em momentos de reflexão em que o estudante avalia a relevância dos e se os relacionados representados fazem sentido, o que contribui na compreensão textual. A visualização dos nodos e arestas de relacionamento foi alterada, tornando-as mais fracas quando não são diretamente conectadas ao nodo selecionado. Esta característica facilita a análise e visualização das informações referentes a cada um dos termos extraídos, sem a necessidade de reorganização do grafo (Figura 9).

Uma característica importante presente na representação das informações extraídas é a capacidade de mostrar ao usuário o número de ocorrências de cada termo e as frases onde este termo é relacionado. Esta característica tem por objetivo permitir que o usuário obtenha informações referentes a termos específicos, tais como frases relacionadas, significado, qual seu papel no texto e quão relevante é em relação aos demais termos apresentados. Além disso, é possível observar o contexto no qual cada termo aparece no texto, promovendo assim a melhor compreensão do termo e de suas relações.

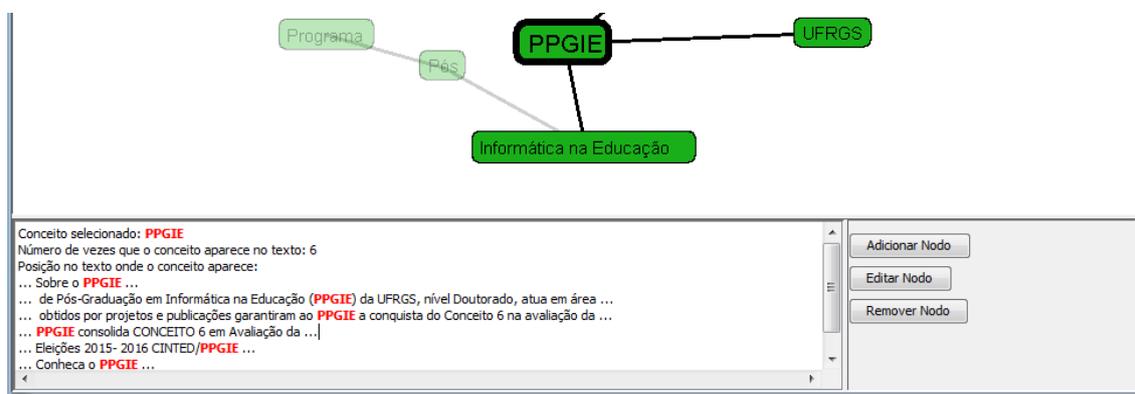


Figura 9 Apresentação das informações do conceito “PPGIE” (imagem gerada pelo autor)

### 5.1.2 Funcionamento do Sobek

O algoritmo responsável pelo minerador Sobek foi aprimorado e alterado de forma significativa durante este trabalho de tese. O seu funcionamento pode ser dividido em três estágios. O primeiro consiste em identificar os termos mais relevantes do texto e

organizá-los por ordem de relevância. No caso do minerador de texto Sobek, um termo será considerado tão importante quanto o número de vezes que ele aparecer no texto (chamado de *frequência* de um termo). O segundo estágio é a criação de relações entre os termos considerados relevantes. O terceiro e último estágio consiste na apresentação do grafo de termos e relações. A seguir, cada um dos estágios será detalhado.

#### 5.1.2.1 Identificação dos termos mais relevantes do texto

O primeiro passo para a identificação dos termos relevantes do texto é separar o texto em palavras. O Sobek utiliza caracteres de espaços em branco, pontuação e sinais de marcação de texto para isso. As palavras são mapeadas em termos, que podem ser *termos simples* (termos que contém apenas uma palavra) ou *termos compostos* (termos que contém duas ou mais palavras). Esse mapeamento de palavras em termos é realizado através de um processo estatístico, que verifica a frequência com que cada palavra e conjunto de palavras são encontrados no texto. Quando um conjunto de palavras aparece constantemente em sequência, é possível que a ideia associada ao conjunto de palavras não possa ser identificada pela soma de termos simples, sendo assim criado um termo composto (Schenker, 2003). Por exemplo, o termo composto “*meio ambiente*” não possui o mesmo significado que a junção dos termos simples “*meio*” e “*ambiente*”.

Para identificar se um termo é simples ou composto, cada palavra é combinada com ‘ $n$ ’ palavras subsequentes, criando termos de tamanho de 1 até  $n$ . Por exemplo, a sequência de palavras ‘*AA BB CC*’ em um cenário onde  $n=3$  irá criar a seguinte sequência de termos: {‘*AA*’, ‘*AA BB*’, ‘*AA BB CC*’, ‘*BB*’, ‘*BB CC*’, ‘*CC*’}. Uma vez que este processo for concluído para todo o texto, aqueles termos com maior frequência são selecionados e os demais são descartados. O valor de  $n$  pode ser maior que 3, mas é pouco provável que no texto haja um termos de mais de 3 palavras que não possa ser representado por termos de menor tamanho. Outro fator importante a considerar é que um valor maior que 3 para  $n$  representaria um considerável acréscimo no custo computacional da operação, tornando-a inviável para valores de  $n$  muito elevados.

As *palavras* que pertencem a termos compostos são removidas da lista de palavras do texto nas ocasiões em que aparecem como um termo composto; todas as demais palavras remanescentes são consideradas termos simples. Isso significa que uma

palavra pode aparecer tanto em um termo composto quanto em um termo simples simultaneamente. Nesse caso, o Sobek identificou que a palavra pertence a um termo composto, mas também é importante para a compreensão do texto que ela seja apresentada como um termo simples.

Durante o processo de identificação de termos, um grupo de palavras chamadas de “*stop words*” é utilizado para remover aquelas palavras que não adicionam informação ao usuário e não são relevantes ao entendimento do texto. Este conjunto está basicamente formado por artigo e proposições. No que concerne o apoio à leitura, objetivo desta tese, a remoção de stop words permite eliminar um grande conjunto de palavras que não contribuem com a compreensão do texto.

Após identificar todos os termos possíveis de serem extraídos do texto, um processo de “*stemming*” é utilizado para remover redundância, termos iguais conjugados em tempos verbais diferentes ou termos que sejam diferentes apenas por serem plural e singular de uma mesma palavra. Este processo junta os termos com grafia diferente mas com mesmo radical em um só termo, evitando que a representação para o usuário esteja poluída com informação redundante ou irrelevante. Por fim, um dicionário é utilizado para remover termos com o mesmo significado semântico. Por exemplo, os termos de “*nação*”, “*país*” e “*pátria*” seriam agrupados em um mesmo termo, já que não há acréscimo em informação ou auxílio ao usuário em deixar os três termos representados no grafo em detrimento de outros termos que possam adicionar diferentes interpretações do texto.

A última parte deste estágio consiste em encontrar um número determinado de termos a serem apresentados. Aqueles termos que têm maior frequência são identificados e armazenados para serem apresentados ao usuário; os demais termos são descartados. Este processo é similar ao processo de *Redução de Dimensionalidade para Representação em Vetor*, proposto por Schenker e colegas (2002). O número de termos relacionados pode ser decidido pelo usuário, mas de acordo com Novak (2002) em média 15 termos são suficientes para representar o tema central de um texto. Neste intuito, a configuração padrão do Sobek é de apresentar aproximadamente 15 termos, representando as informações que aparecem com maior frequência no texto. Há diversas formas de alterar o número de termos extraídos, mas é importante ressaltar que um número maior de termos pode prejudicar a visualização e identificação daqueles mais

relevantes, enquanto um número menor de termos pode não fornecer informações suficientes para a boa representação do texto.

### 5.1.2.2 Relacionamento entre termos

O segundo estágio no processo de mineração de texto é o de identificação dos relacionamentos entre os termos. Um relacionamento entre termos é formado quando eles estão próximos um do outro no texto. Isso pode representar diversos tipos de informação acerca dos dois termos, tais como relação de causa e consequência, uma relação temporal ou mesmo que os termos têm significados relativos um ao outro. A análise do texto relaciona dois termos quando estes estão distantes não mais que 'z' termos um do outro e quando não há marcação de final de parágrafo entre eles (Schenker, 2003).

Cada termo pode estar relacionado com mais de um, o que poderia produzir um grafo representativo do texto, mas com muitas relações e difícil de compreender. Para reduzir o número de relações que um termo pode ter e apresentar apenas as mais relevantes (tornando assim o grafo mais claro, expressivo e acessível ao usuário), um número máximo de 'r' relações é permitido para cada termo. O número exato de ligações que um termo 'c' pode ter depende de sua frequência em relação à frequência do termo mais recorrente no texto. Isso permite que termos mais recorrentes (e, teoricamente, mais relevantes ao texto) possuam mais conexões que os demais termos e se destaquem em posições centrais do grafo. Assim, o número de relações de um termo é equivalente à '*r multiplicado pela sua frequência e dividido pela frequência do termo mais recorrente*'. Se um termo tiver uma lista de relacionamentos maior que o máximo de relações permitidas, apenas aqueles relacionamentos que aparecem mais frequentemente são apresentados no grafo, sendo os demais descartados. Não há um número mínimo de relações que um termo pode ter, podendo até mesmo não se relacionar com nenhum outro termo do grafo (equivalente a ter zero relação).

Os valores de 'z' e 'r' são pré-estabelecidos no Sobek, não sendo possível serem alterados pelo usuário. Sobek usa valores de 'z = 5' e 'r = 7'. Esses valores foram estabelecidos com base em avaliações de usuários e análise de dados. Um valor de 'z' maior que cinco costuma apresentar relações de pouca importância entre termos; um valor de 'r' maior que sete pode produzir grafos com um número demasiadamente alto

de relações, podendo até mesmo produzir grafos onde todos os termos se relacionam entre si.

### 5.1.2.3 Visualização do grafo de termos

O estágio final do Sobek é a construção do grafo utilizando as informações extraídas do texto. Para representar o grafo visualmente, é utilizado a *Prefuse API*<sup>5</sup>. No grafo, os termos são representados como vértices (*nodos*) e as relações entre eles são representados como conexões (*arestas*). Para melhorar a visualização do grafo, cada nodo tem uma cor e um tamanho diferente, baseado na frequência do termo que ele representa. Quanto maior e mais escuro for o nodo (variando em tons de verde), maior a frequência de ocorrência desse termo no texto. A frequência de um nodo também influenciará no seu número de ligações, tornando os nodos com maior frequência centrais no grafo, sendo estes mais conectados e destacados dos demais.

A interação do usuário com o grafo é parte fundamental da ferramenta Sobek. O minerador busca criar um grafo que seja o mais representativo possível do texto minerado, mas isso nem sempre é possível. Diversos fatores podem dificultar a representação do texto pelo Sobek, tais como o gênero narrativo do texto, a quantidade de palavras presentes nele ou mesmo se o texto não tiver um foco central bem definido.

O grafo resultante da mineração tem uma série de funcionalidades que permitem ao usuário personalizar ele ou mesmo gerar grafos inteiramente novos. É possível adicionar e remover nodos e conexões, assim como interagir com os nodos e mudar eles de posição ou cor. Outra importante funcionalidade do grafo é a capacidade de selecionar nodos e receber informações específicas sobre determinado termo, tais como posições no texto onde aparece e a sua frequência no texto. Esta interação do usuário com o grafo visa incentivar a reflexão dele acerca das informações sobre as quais o usuário está interagindo. Ademais, as informações acerca dos termos e suas relações têm por objetivo auxiliar a promover a aprendizagem significativa, conforme descrito na sessão 5.2 deste trabalho.

Do ponto de vista pedagógico, esta interação vai além de simplesmente melhorar a representatividade do texto. Ao interagir com o grafo, o usuário está refletindo sobre o

---

<sup>5</sup> <http://prefuse.org/>

que foi lido no texto e sobre o papel que cada termo tem na representação do texto. A remoção de um nodo, por exemplo, representa a noção do usuário de que aquele termo contido no nodo não faz parte da representação do texto. Isso reflete um pensamento crítico acerca do texto e de seu significado, instigando a pessoa a expor as suas ideias e a sua percepção do texto em forma de um grafo.

A interação do usuário junto ao grafo (seja para adicionar, remover ou editar parte dele) permite também adequar o grafo à experiência do usuário, personalizando o conhecimento e tornando o aluno parte essencial do minerador Sobek. Ao contrário de uma simples exposição de conhecimento ao aluno, esta funcionalidade é essencial para tornar ele parte ativa e central no processo de educação. Alunos com bases de conhecimento diferentes manipulam o grafo de forma diferente (criando uma rede de informações próprias) ao mesmo tempo em que instiga o usuário a fazer reflexões sobre o que entende acerca de texto e como relacionar esses conceitos com os de conhecimento prévio.

### **5.1.3 Validação da ferramenta de mineração Sobek**

Antes de utilizar o minerador Sobek para realizar os experimentos propostos neste trabalho e descritos no final deste capítulo, foram realizadas atividades que visam prover uma validação da ferramenta de mineração de texto quanto à finalidade para a qual se destina (extração de termos e representação de textos em formatos de grafos). Para esta validação foi requisitado o auxílio de especialistas na área de educação (estudante de doutorado e professores do curso de Pós Graduação em Informática na Educação) e foram utilizados textos referentes às áreas de especialização dos participantes. No total, 2 textos foram utilizados e 6 pessoas participaram desta avaliação. Todos os textos e a ferramenta de mineração Sobek estão disponíveis online e os resultados dos testes de validação podem ser visualizados através do site [http://sobek.ufrgs.br/Resultados\\_2015](http://sobek.ufrgs.br/Resultados_2015).

A validação da ferramenta procedeu em duas etapas:

- Comparação da lista de termos extraídos pelo Sobek com a lista fornecida por especialistas.
- Validação da lista de termos extraídos pelo Sobek;

A seguir está detalhado como procedeu cada etapa de validação e os resultados obtidos.

### **5.1.3.1 Comparação entre a lista de termos extraídos pelo Sobek e a lista de termos fornecidos por especialistas**

Para avaliar a *qualidade* das informações extraídas pelo Sobek (sendo a metodologia utilizada a contestação de que os termos extraídos de fato representam o conteúdo do texto), foi utilizado um teste do tipo “*Gold Standard*”. Em estatística, o *gold standard* se refere ao resultado de um teste (ou referência) acerca da melhor avaliação obtida dentro das condições possíveis. Por exemplo, ao fazer a medição de poluentes da água de um rio, um teste do tipo gold standard vai detectar (da melhor forma possível) os componentes químicos existentes na água com base nas condições de experimento (horário, química, local, maquinário disponível, etc.). Não necessariamente o resultado do teste será perfeito, mas ele é considerado válido para futuras referências acerca dos poluentes contidos na água daquele rio.

Ao validarmos uma ferramenta de extração de conceitos, além de avaliar se o resultado apresentado por ela é satisfatório também gostaríamos de investigar o quão próximo de um *gold standard* ele está; ou seja, qual a *qualidade* da informação retornada pelo Sobek. Sendo utilizados textos referentes a áreas educacionais, foi considerado que professores e doutorandos dessa área devem ser capazes de identificar nesses textos os termos mais relevantes. Assim, foi possível construir um gold standard do que seriam os termos principais de cada texto, utilizando este conjunto de termos extraídos pelos especialistas. Essa lista de gold standard serviu de métrica para a avaliação dos termos extraídos pelo Sobek, a fim de validar o resultado da extração de termos não apenas em satisfatória ou não satisfatória, mas também indicando o grau de precisão e concordância entre os termos presentes na lista gold standard e os termos extraídos pelo Sobek.

No intuito de diminuir erros estatísticos, foi considerada a lista de termos fornecida por três especialistas para cada texto analisado e cada termo foi incluído na lista *gold standard* se estava presente na lista de pelo menos dois dos três especialistas. Uma vez que a escolha dos termos era livre, cada especialista escreveu o termo como

considerou mais propício. Assim, uma análise prévia à comparação necessitou ser realizada para averiguar se os termos possuíam a mesma definição e significado, mesmo possuindo grafias diferentes.

Os quadros a seguir (quadros 1 e 2) apresentam as listas de termos produzidas por cada especialistas para o primeiro texto analisado e a comparação entre a lista considerada “*Gold Standard*” e a lista produzida pelo Sobek.

**Quadro 1: Lista de termos considerados relevantes pelos especialistas para o *primeiro* artigo**

Termo	Especialista A	Especialista B	Especialista C
1.	Conta	Facebook	Facebook
2.	Criança	Sala de Aula	Comunicação
3.	Sala de Aula	Comunicação	Famílias
4.	Ferramenta de Comunicação	Aprendizagem	Aprendizagem
5.	Facebook	Pais	Aprendizagem diária
6.	Aprendizagem	Pais	On-line
7.	Pais	Escola	Plataforma
8.	Escola	Mídia Social	Escola
9.	Mídia Social	Estudantes	Mídia Social
10.	Estudantes	Professores	Estudantes

**Quadro 2: Lista de termos que compõe a lista gold standard, que são considerados relevantes pelo Sobek e se ambas listas concordam com a classificação (para o primeiro artigo).**

Termo	Lista Gold Standard	Sobek	Concordam?
1.	Sala de aula	Sala de aula	Sim
2.	Comunicação	Comunicação	Sim
3.	Facebook	Facebook	Sim
4.	Aprendizagem	Aprendizagem	Sim
5.	Pais	Pais	Sim
6.	Escola	Escola	Sim
7.	Mídia Social	Mídia Social	Sim
8.	Estudantes	Estudantes	Sim
9.		Benefícios	Não
10.		Sherry	Não

Através do teste de gold standard é possível calcular a medida de *sensibilidade* (também conhecida como *verdadeiro positivo* ou *recall*) na qual os termos considerados relevantes pelo Sobek ( $nS$ ) e pelos especialistas ( $nE$ ) são levados em consideração. Para o primeiro artigo, a *sensibilidade* calculada para a capacidade do Sobek de identificar um termo relevante é dada por:

$$sensibilidade = nS / nE = 8 / 8 = 1.0$$

Outra medida que pode ser calculada a partir das listas de termos é a *precisão* das listas. O valor de *precisão* é calculado dividindo o número de termos corretamente identificado pelo Sobek (ou seja, aqueles que aparecem na lista de gold standard) pelo número total de termos apresentados pelo Sobek ( $tS$ ). Para o primeiro artigo, esse valor é de:

$$precisão = nS / tS = 8 / 10 = 0.8$$

O valor encontrado pelo cálculo de *sensibilidade* indica que todos os termos apontados como relevantes pelo Sobek também foram considerados relevantes por pelo menos dois de três especialistas. Isso indica que o Sobek tem grande capacidade de encontrar os termos que são considerados chaves para o texto e para a sua compreensão. A *precisão*, cujo valor foi de 80%, serve de indicador para afirmar que não apenas o

Sobek encontrou os termos relevantes como também o fez sem apresentar demasiadas informações consideradas irrelevantes pelos especialistas.

Um segundo texto foi lido e analisado por 3 especialistas diferentes daqueles que participaram da avaliação do primeiro texto. As listas de termos produzidas por cada especialista e a comparação entre a lista considerada “*Gold Standard*” e a lista produzida pelo Sobek estão apresentados a seguir (quadros 3 e 4).

**Quadro 3: Lista de termos considerados relevantes pelos especialistas para o *segundo* artigo**

Termo	Especialista A	Especialista B	Especialista C
1.	Benefícios	Acesso expandido à Wi-Fi	Ônibus
2.	Ônibus	Acesso à Internet	Acesso à Internet
3.	Internet	Ambiente de aprendizagem	Aprendizagem
4.	Aprendizagem	Oportunidades para aprendizagem	Móvel
5.	Acesso Móvel	4G Padrão	Rede
6.	Rede	Educação baseada na tecnologia	Oportunidades
7.	Escola	Acesso sem fio	Escola
8.	Estudantes	Rede sem fio	Tecnologia
9.	Tecnologia na Educação	Wi-Fi	Rede sem Fio
10.	Rede sem Fio		wi-fi

**Quadro 4: Lista de termos que compõe a lista gold standard, que são considerados relevantes pelo Sobek e se ambas listas concordam com a classificação (para o *segundo* artigo).**

Termo	Lista Gold Standard	Sobek	Concordam?
1.	Ônibus	Ônibus	Sim
2.	Internet	Internet	Sim
3.	Aprendizagem	Aprendizagem	Sim
4.	Rede	Rede	Sim
5.	Escola	Escola	Sim
6.	Wi-Fi	Wi-Fi	Sim
7.	Rede sem fio	Rede sem fio	Sim
8.	Oportunidades	Móvel	Não
9.		Estudantes	Não
10.	Tecnologia na Educação	Tecnologia	?

Os mesmos valores de *sensibilidade* e *precisão* podem ser calculados para o segundo artigo analisado:

$$sensibilidade = 8 / 9 = 0.88$$

$$precisão = 8 / 10 = 0.8$$

Novamente foram encontrados valores bastante altos tanto para a *sensibilidade* quanto para *precisão*, reforçando os resultados obtidos no experimento anterior.

Considerando os valores de *sensibilidade* e *precisão*, é possível obter o valor do *F-Score* (chamado também de *F-measure*) para cada um dos artigos. Este valor é considerado como sendo a acurácia do teste realizado e leva em consideração os dois valores previamente calculados, fornecendo uma média ponderada dos valores. No caso de ambos os valores serem considerados igualmente importantes, a mediação é denominada de  $F_1$ -score.

O valor de  $F_1$ -score é calculado por:

$$F_1 = 2 * \frac{precisão * sensibilidade}{precisão + sensibilidade}$$

De acordo com a equação, o valor de  $F_1$  para o primeiro artigo ( $F_{1A}$ ) e para o segundo artigo ( $F_{1B}$ ) são:

$$F_{1A} = 2 * \frac{0.8 * 1.0}{0.8 + 1.0} = 0.89 \quad F_{1B} = 2 * \frac{0.88 * 0.8}{0.88 + 0.8} = 0.84$$

Tais valores permitem concluir que o minerador Sobek não apenas é capaz de encontrar termos considerados chaves para o texto mas também o faz sem adicionar informações irrelevantes ao resultado da mineração. O resultado próximo de 1 confirma que há um bom balanço entre a *sensibilidade* e a *precisão*, demonstrando a boa capacidade do Sobek encontrar exatamente os termos considerados relevantes no texto.

### 5.1.3.2 Validação da lista de termos extraídos pelo Sobek

O índice Kappa (também chamado de *Coefficiente Kappa de Cohen*) é uma medida estatística de concordância entre avaliadores para itens qualitativos (Carletta, 1996). Ele mede o grau de concordância entre dois ou mais avaliadores ao classificarem itens e desconsidera a concordância que seria esperada de tão somente acaso. É geralmente utilizado por ser uma medida mais robusta do que o simples cálculo de por cento de concordância, uma vez que leva em consideração que o acordo pode ocorrer por acaso (de forma aleatória). O índice Kappa mede o acordo entre dois avaliadores que classificam  $N$  itens em  $C$  categorias mutuamente exclusivas. O resultado é um valor de concordância entre 0 e 1, onde 1 representa que há total concordância entre os especialistas e o Sobek e o valor 0 indica que não houve concordância entre eles ou que a concordância se deu por sorte.

Apesar do coeficiente de Kappa ser amplamente aceito no ambiente científico, alguns pesquisadores expressam preocupação na tendência do organizar os dados em categorias pré-definidas, subestimando a possibilidade de utilizar outras categorias que poderiam ser utilizadas (Strijbos et. al. 2006). Por isso, o índice Kappa é considerado uma medição bastante conservadora quando se busca identificar a concordância entre avaliadores. Porém, no âmbito desta avaliação, há apenas duas possibilidades de classificação de cada termo (relevante ao texto ou não) e isso favorece a validação da ferramenta através do cálculo do índice Kappa.

O valor do coeficiente Kappa ( $K$ ) é definido por:

$$K = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)},$$

sendo  $\text{Pr}(a)$  a concordância relativa dos observadores e  $\text{Pr}(e)$  a probabilidade hipotética de concordância aleatória (o acaso).

O coeficiente Kappa foi utilizado para calcular o grau de concordância entre a lista de termos provida pelo Sobek e aquela considerada a gold standard. Cada termo do texto foi classificado como sendo um termo chave do texto (chamado de *relevante*) ou um conceito não-chave (chamado de *irrelevante*). Uma vez que os termos só podem estar presentes em uma lista ou em outra, todos os termos que aparecem no texto e que não foram considerados explicitamente relevantes (pelo Sobek ou pela lista gold standard) foram automaticamente classificados como *irrelevantes*.

Nesta avaliação, não foi possível considerar termos compostos por dois motivos:

- O número de termos *irrelevantes* seria demasiadamente grande, pois todos os conceitos simples deveriam ser também organizados em conceitos compostos e isso elevaria de forma exponencial o número de termos *irrelevantes*.
- A forma de escolha do número máximo de conceitos simples que devem ser agrupados para formar um conceito composto acrescentaria um viés na avaliação que pode ser contestado.

O primeiro artigo analisado possui um total de 453 termos simples. A lista de termos extraída pelo Sobek possui 10 termos relevantes, considerando assim 443 termos como *irrelevantes*. A lista considerada gold standard possui 8 termos relevantes, considerando 445 termos como irrelevantes. As listas possuem 7 termos em comum.

O número de termos classificados de forma igual por ambas as listas soma 451 termos, enquanto que a quantidade de termos considerados iguais que se espera de um resultado aleatório é de 435 termos. Com base nesses valores, podemos calcular que o *coeficiente de Kappa* para este primeiro artigo é de 0,79, o que é considerado um valor onde há uma *concordância substancial* entre os observadores. Este valor é considerado alto para o coeficiente Kappa, indicando assim que a similaridade entre as listas do Sobek e de Gold Standard não ocorreu pelo acaso.

Considerando as mesmas métricas para o segundo artigo, que contém 321 termos simples, vemos que a lista do Sobek possui 10 termos e a lista de gold standard

9, sendo 8 deles considerados relevantes por ambas medições. Neste cenário, o número total de termos igualmente classificados pelas duas listas foi de 318, enquanto que a quantidade esperada considerando o fator aleatório é de 302 termos. Como resultado, o *coeficiente de Kappa* para este segundo artigo atinge o valor de 0,837, o que é considerado como uma *concordância quase perfeita* entre os observadores. Estes resultados demonstram o bom nível de precisão da ferramenta no que diz respeito ao seu desempenho na extração de conceitos considerados relevantes no texto.

## **5.2 O minerador de textos Sobek como apoio à compreensão textual**

O minerador de textos Sobek foi desenvolvido, inicialmente, com o intuito de auxiliar professores e estudantes de cursos de ensino a distância, tendo em vista a extensa quantidade de material escrito que provém deste tipo de ensino. Uma vez que o Sobek foi idealizado como uma ferramenta de apoio pedagógico, seus aspectos tanto do ponto de vista da usabilidade quanto da interação com o usuário foram desenvolvidos com base em teorias educacionais. A interface do Sobek foi projetada para que fosse simples de utilizar, não sendo necessário nenhum treinamento prévio. Apesar de simples de utilizar, ela possui uma ampla gama de configurações, possibilitando a personalização da ferramenta para diferentes fins e ambientes. Esta característica torna o Sobek propício para o uso em atividades básicas de letramento ou atividades complexas de análise de textos.

Diversas pesquisas foram levadas em consideração durante o planejamento da ferramenta, em especial as pesquisas de Ausubel (1968) e sua teoria de aprendizagem significativa. A base da teoria de aprendizagem significativa de Ausubel é que o aprendizado ocorre ao se assimilar novos conceitos e proposições no corpo de conhecimento que a pessoa já possui. A aprendizagem dita significativa ocorre quando um conhecimento novo é assimilado pela pessoa e conectado a informações já conhecidas por ela. Caso a informação não seja integrada de forma significativa ao conhecimento já consolidado do aprendiz, o aprendizado é chamado de *mecânico*, que ocorre quando um fato é percebido e pelo indivíduo mas não é associado às informações já conhecidas por ele.

A aprendizagem significativa é um processo complexo que necessita de três condições para ocorrer:

- O conceito (ou ideia) a ser aprendido deve ser claro e apresentado com linguagem e exemplos relacionáveis a conceitos previamente conhecidos pelo indivíduo.

- O indivíduo deve possuir um conhecimento extenso acerca do assunto que se está sendo aprendido.

- O indivíduo deve querer aprender de forma significativa esse novo conceito (ou ideia), tanto através da predisposição afetiva quanto cognitiva. Conforme comentado no Capítulo 3, Novak e Cañas (2007) argumentam que essa condição pode apenas ser apenas indiretamente influenciada pelo mentor ou professor, incentivando os estudantes a buscarem aprender novas informações ao invés de simplesmente decorá-las, o que seria uma aprendizagem mecânica das informações. Nesse cenário, é importante motivar os estudantes para que eles queiram aprender e incentivar o pensamento crítico, fomentando assim o relacionamento de conteúdos novos aos já conhecidos e desenvolvendo a aprendizagem significativa.

Apesar de a aprendizagem significativa proporcionar uma forma de aprendizado considerada mais completa que a aprendizagem mecânica, não significa que a última não possibilite ao indivíduo registrar informações na memória de longa duração. A diferença entre as duas formas de aprendizagem está relacionadas à forma de integração delas na estrutura cognitiva da pessoa. Uma vez que a aprendizagem mecânica não relaciona novas informações ao conhecimento já construído pelo indivíduo, estas novas informações tendem a ser esquecidas se não forem reforçadas constantemente. Outra consequência negativa dessa forma de aprendizagem é que as informações aprendidas não são refinadas ou modificadas ao longo do tempo, tornando difícil modificá-las ou corrigir informações incorretas. Equívocos e informações falsas podem persistir na memória sem que novas informações possam alterar/influenciar as anteriores (Novak, 2002).

Baseado nestas teorias, o minerador de textos tem por objetivo auxiliar a compreensão de textos e fomentar a aprendizagem significativa. O diferencial do Sobek em relação a outros mineradores de texto está na forma de representação das informações extraídas do texto. A representação em forma de grafo busca aproximar-se de um mapa conceitual simplificado. Nesta representação, os principais conceitos do texto aparecem de forma diferenciada uns dos outros, sendo os conceitos considerados

mais relevantes representados por nodos maiores com um maior número de conexões. Também estão presentes no grafo relações entre os conceitos que buscam auxiliar o aluno a relacionar as informações contidas no texto. Através da visualização dos conceitos e de suas relações, objetiva-se que o leitor consiga identificar os conceitos considerados fundamentais no texto e associá-los a seu próprio conhecimento e a outras ideias do próprio texto. As conexões entre os conceitos foram projetadas para que o leitor possa compreender melhor como cada conceito está inserido no contexto do texto.

Para que fosse possível apresentar um grafo contendo uma quantidade de informações considerada suficiente para a representação do texto mas que não sobrecarregasse o estudante de informações (um grafo nem demasiadamente grande que o sujeito não consiga compreender e nem muito pequeno que omita dados relevantes do texto). O número de conceitos apresentados no grafo é baseado nas teorias de Miller (1965) e Novak (2002). De acordo com Miller, a memória do ser humano é um complexo sistema de memórias inter-relacionadas. Apesar de todas as memórias serem interdependentes, um componente importante da memória que permite levar o conhecimento à *memória de longo prazo* é a chamada *memória de curta duração*. A premissa do autor é que a memória de curta duração só pode processar um número limitado de informações não relacionadas de forma simultânea (entre cinco e nove itens diferentes), não sendo produtivo apresentar a um indivíduo um número de itens maior que este em um determinado instante (o autor descreve essas informações como sendo unidades psicológicas, ou em inglês *psychological units*).

Apesar da memória de curta duração poder processar uma quantidade limitada de informação, quando esta informação é relacionada entre si a capacidade de processamento da memória humana é elevada. Por exemplo, se uma pessoa precisa se lembrar de diversas palavras não relacionadas, ela provavelmente se lembrará de apenas 5 a 9 palavras. Porém, se as palavras contém relação umas com as outras, a pessoa poderá lembrar-se de mais de 12 palavras facilmente (Novak 2002). O mesmo princípio se aplica à relação entre essas palavras. Não é possível lembrar-se de todas as relações entre as palavras, apenas de algumas.

Com base nessas pesquisas, o Sobek foi configurado para apresentar um número limitado de conceitos relacionados a cada texto (aproximadamente 15 conceitos) e um número menor de relações entre os conceitos (não mais de 7 relacionamentos para um conceito). O valor aproximado de 15 conceitos está de acordo com as teorias de Novak (2002) para a construção de mapas conceituais. De acordo com o autor, a partir do

contexto e área aos quais o texto se relaciona, o número de conceitos presentes para a construção do mapa conceitual deve ser entre 15 e 25. Apesar de o número de conceitos ser maior que aquele considerado ideal para o processamento da memória de curta duração, isso permite que o usuário explore o grafo e os conceitos de maneira individual, não induzindo todos os leitores a terem a mesma visão sobre o texto. Alguns leitores podem analisar um determinado aspecto do grafo enquanto outros leitores podem ter uma visão diferente. Apesar de permitir essa individualidade na visão do grafo, vale ressaltar que os conceitos considerados mais relevantes são apresentados em tamanho maior e de forma mais visível que os demais, buscando ressaltá-los e ao mesmo tempo possibilitar a compreensão do grafo por diferentes ângulos. Também é importante ressaltar que os conceitos estão presentes de forma escrita dentro de nodos que compõem o grafo. A forma escrita de representação dos conceitos tem por objetivo estimular o reconhecimento visual destes conceitos pelo leitor e auxiliá-los no reconhecimento dos mesmos conceitos dentro do texto.

A interação do usuário com o grafo é outro importante aspecto voltado a auxiliar a compreensão textual e promover a aprendizagem significativa dos conceitos relevantes do texto. Esta interação ocorre inicialmente quando o usuário reorganiza os nodos que contém os conceitos extraídos no grafo. Essa organização, quando realizada pelo usuário, o incentiva a pensar sobre os conceitos, suas relações e a importância de cada um no texto. O usuário organizará o grafo de forma a poder visualizar melhor algum aspecto que ele considere relevante do texto, auxiliando-o a refletir e compreender melhor algum aspecto. Juntamente com a organização do grafo, o usuário pode adicionar, remover ou editar os nodos. Esse processo o auxilia na organização de ideias e no relacionamento das informações. Uma vez que o leitor está editando os nodos do grafo, ele está refletindo sobre os conceitos e interagindo com cada um em particular, processo que apoia a aprendizagem significativa. A interação com o grafo (edição, remoção e adição de nodos/ligações) faz com que o estudante se posicione como sujeito ativo, resgatando em sua memória o que sabe sobre o assunto, contrastando estas informações com o que está representado no grafo, e fazendo os ajustes que julgar necessários. Neste caso, nota-se claramente o fator do indivíduo de '*querer aprender*', fator esse necessário para que a aprendizagem significativa ocorra. Ao editar o conceito, o usuário está informando que o grafo pode ser mais representativo do que é, sendo este um processo que só pode ser realizado pela

compreensão das ideias do texto e pela reflexão acerca das informações representadas no grafo.

Por fim, a última forma de interação com o grafo ocorre quando o estudante seleciona um nodo. Ao selecionar o nodo, os conceitos que possuem relações diretas com o nodo selecionado permanecem visíveis ao usuário, enquanto os demais nodos e relações tem sua cor clareada. Isso auxilia o usuário a visualizar as relações de cada conceito, permitindo identificar os conceitos que possuem informações relevantes acerca de um determinado conceito extraído do texto. Essa estratégia também auxilia o relacionamento do conceito selecionado a outras informações sobre as quais o usuário pode já ter conhecimento, facilitando assim a assimilação dele. Outra ação que ocorre ao selecionar um nodo é a exibição de informações complementares acerca desse nodo, tais como número de ocorrência dele no texto e passagens do texto onde ele aparece. Assim, o leitor pode identificar os trechos do texto onde o conceito é relacionado, podendo reler tais trechos e melhor compreender o texto.

### **5.3 Avaliação da Ferramenta Sobek no Apoio à Compreensão Textual**

A pesquisa desta tese foi delineada com método de coleta e análise de dados quantitativo e qualitativo. A investigação foi desenvolvida em duas etapas. Na primeira (chamada de pesquisa preliminar), foram realizadas intervenções junto a um pequeno grupo de alunos de uma escola municipal de Porto Alegre. Estas intervenções focaram em diferentes aspectos do letramento e em como o Sobek poderia auxiliar no desenvolvimento de produção e compreensão textual. Após esta etapa, a ferramenta e os métodos de intervenções foram refinados, sendo especificado o enfoque do uso do minerador de textos para compreensão textual. Uma segunda série de intervenções foi realizada em uma escola particular de Caxias do Sul, com um número maior de estudantes. O escopo de ambas as atividades está descritos a seguir.

### **5.3.1 Pesquisa preliminar**

Durante a etapa inicial do desenvolvimento desta tese, foram realizadas intervenções com o intuito de avaliar as implicações da utilização do Sobek para as atividades compreensão textual. No desenvolvimento das atividades, também foram incluídas tarefas de produção textual, como forma de apoio ao letramento. Os resultados aqui apresentados, contudo, terão como foco as atividades de leitura e compreensão textual.

#### **5.3.2.1 Sujeitos da pesquisa**

A pesquisa foi realizada com um grupo de estudo de alunos do sétimo, oitavo e nono ano do Ensino Fundamental da Escola Municipal Lidovino Fanton, de Porto Alegre, e ocorreu entre 08/10/2015 e 26/11/2015. Esse grupo foi formado por estudantes que apresentavam deficiências de leitura e produção textual e recebiam reforço escolar com atividades relacionadas ao letramento. A participação desses estudantes na pesquisa foi voluntária, sendo apenas sugerida por seus professores com base em suas produções em sala de aula. É importante ressaltar que em todas as intervenções e atividades realizadas foi explicitado que os estudantes poderiam desistir de participar da pesquisa a qualquer momento, sem nenhum tipo de consequência a eles ou à sua atuação no grupo de estudos.

Um total de 20 alunos iniciou a pesquisa, sendo divididos em duas turmas de 10 alunos cada. As turmas continham alunos de diferentes séries, sendo separados de acordo com a discricção da professora responsável pela atividade de reforço escolar dos estudantes.

Os responsáveis legais pelos estudantes participantes da pesquisa assinaram um termo de consentimento permitindo que eles participassem das atividades. Este termo se encontra no Apêndice A. A diretora da escola Lidovino Fanton também assinou um termo de consentimento para que as atividades pudessem ser realizadas na escola (Apêndice B).

#### **5.3.2.2 Materiais das intervenções**

Durante as intervenções, foram utilizadas imagens retiradas de sites da Internet (Figura 10 e Figura 11) com o intuito de instigarem discussão entre os alunos e

propiciarem argumentos para a produção textual. Estas imagens foram escolhidas de forma arbitrária pelo pesquisador e apresentam locais conhecidos pelos estudantes e atividades realizadas de forma corriqueira.

Para as intervenções focadas em compreensão textual, foram utilizados textos da ANA (Avaliação Nacional de Alfabetização). Estes textos fazem parte da avaliação voltada aos estudantes de quinto ano e oitavo ano.



**Figura 10** Imagem da praia de Ipanema em Porto Alegre.



Figura 11 Jovem andando de skate pelo “calçadão” da praia de Ipanema, Porto Alegre.

### 5.3.2.3 Descrição do experimento

As intervenções ocorreram durante cinco encontros, com intervalos de uma ou duas semanas entre cada uma. Cada encontro teve um tempo médio de 2 horas de duração e possuía objetivos específicos a serem desenvolvidos.

Durante o primeiro encontro, a intervenção realizada teve dois objetivos:

1. Estabelecer uma comunicação entre os estudantes e os pesquisadores, permitindo avaliar a estrutura disponibilizada para a atividade e as características dos estudantes;
2. Apresentar a ferramenta de mineração de textos Sobek aos alunos, demonstrando seu uso e potencial.

Nessa atividade, foram realizadas as apresentações formais, explicando as questões relacionadas à pesquisa e ao que seria desenvolvido naquela atividade e nas subsequentes. Foi também explicado e distribuído o termo de consentimento. O termo de consentimento foi lido em voz alta e cada aluno levou um para ser assinado pelo seu responsável.

Após a etapa inicial de apresentação, a questão da importância do letramento tanto na vida acadêmica como pessoal de cada um foi abordada. O minerador de textos Sobek foi apresentado, demonstrando o resultado gráfico da extração de conceitos de

um texto e como ele poderia auxiliar na compreensão textual dos estudantes. Diversas simulações de uso da ferramenta Sobek foram realizadas e dúvidas acerca da ferramenta e do projeto foram respondidas.

No segundo encontro, a proposta escolhida teve por objetivo familiarizar os estudantes com a ferramenta e motivar o processo criativo de escrita dos mesmos. Nessa atividade, foi apresentada uma figura para os estudantes e foi debatida uma história fictícia relacionada com a imagem. A finalidade dessa ação foi fomentar a criatividade dos estudantes e demonstrar como poderia ser pensada uma história ou um conjunto de situações com base em uma imagem.

Após esta etapa inicial, uma segunda figura foi apresentada. Os estudantes foram requisitados a produzir um grafo utilizando a ferramenta Sobek, contendo os principais termos de uma história criada por cada um deles, com base na imagem. Nesse momento, os estudantes estavam se habituando à criação e edição de nodos no grafo, bem como identificando como o Sobek poderia auxiliá-los a melhor organizar as suas ideias. Uma vez finalizados os grafos, os estudantes trocaram de computador, sentando a frente de grafos produzidos por colegas. Em uma segunda etapa, cada aluno deveria elaborar um texto com base nos termos e relações presentes no grafo construído por um colega. A identificação das informações apresentadas pelo Sobek também foi exercitada nessa atividade.

Durante o terceiro encontro, o objetivo foi o de observar o processo de compreensão textual dos alunos com o auxílio do Sobek. Na terceira intervenção, foram debatidos os textos criados previamente pelos alunos bem como suas percepções em relação à ferramenta Sobek. Nesse encontro foi demonstrado de forma mais específica como o Sobek poderia auxiliar a compreensão de textos através da criação e interpretação de grafos. Foi fornecido aos estudantes um conjunto de textos, sendo dois textos para os alunos do sétimo ano e outros dois para alunos do oitavo e nono ano. Um questionário com perguntas acerca dos textos foi respondido.

O quarto encontro foi novamente voltada à produção textual, dessa vez partindo de um grafo gerado pela ferramenta Sobek. Cabe aqui enfatizar que, apesar do objetivo desta tese estar relacionado à avaliação da ferramenta Sobek no apoio à leitura e compreensão textual, as atividades de escrita foram introduzidas como forma de dinamizar os encontros e engajar mais os estudantes nas atividades propostas. Durante o

quarto encontro, inicialmente foi apresentado um grafo aos estudantes com base em um texto que versava sobre os efeitos negativos do cigarro no organismo. A partir das informações apresentadas no grafo foi proposta uma discussão, onde os conceitos considerados mais relevantes pelo Sobek e suas conexões serviram de norteadores. Essa discussão serviu para reforçar junto aos estudantes os aspectos fundamentais da leitura de grafos e da identificação de conceitos chaves no texto, que são a base do processo de compreensão textual. Em seguida, foi apresentado um segundo grafo aos alunos, gerado de um texto acerca das consequências de misturar álcool e direção. Com base nesse segundo grafo, cada estudante deveria escrever um texto utilizando os termos destacados e seus relacionamentos. Os estudantes foram instruídos que esse texto deveria possuir todas as informações que eles conseguissem extrair do grafo, além de conclusões do próprio aluno a respeito do assunto. Destaca-se que os estudantes não tiveram acesso aos textos que serviram de base para a geração dos grafos. O objetivo aqui foi oferecer uma maior liberdade para que os alunos desenvolvessem a sua criatividade na escrita dos textos dentro das relações apresentadas pelo grafo.

A quinta e última intervenção ocorreu de maneira mais informal que as demais e teve por objetivo tanto avaliar a compreensão textual dos alunos quanto a sua percepção em relação ao uso do minerador e das atividades realizadas. De forma semelhante ao terceiro encontro, foram passados dois textos aos estudantes, que deveriam ler e responder a um questionário baseado nesses textos. A utilização do Sobek para geração de um grafo a partir dos textos foi opcional. Ao término dessa etapa, os estudantes responderam um questionário acerca do uso da ferramenta, suas percepções e opiniões, tanto sobre o Sobek quanto às atividades realizadas.

#### **5.3.2.4 Resultados da fase preliminar**

As atividades destinadas à avaliação do uso do minerador de textos Sobek para apoio à compreensão textual foram realizadas durante duas intervenções, além da primeira intervenção para demonstração e tutorial sobre o uso da ferramenta. Na primeira das duas intervenções destinada ao assunto, foi fornecido aos estudantes dois textos, ambos retirados da ANA. Os alunos de sexto ano receberam textos referentes à avaliação ANA para quinto ano e os demais alunos receberam textos destinados à avaliação de alfabetização do oitavo ano. A escolha dos textos e a divisão dos estudantes ocorreram com apoio da professora responsável pela ação de apoio escolar, que esteve

presente durante todas as atividades de pesquisa na escola. De acordo com a professora, os textos estavam de acordo com a capacidade esperada pelos estudantes, assim como as questões elencadas para cada texto.

Juntamente com os textos, foram fornecidos questionário com perguntas sobre os textos. Devido ao fato da *ANA* possuir um número demasiado reduzido de questões para cada texto (em geral há apenas duas questões por texto), foi necessário acrescentar perguntas elaboradas especificamente para esta pesquisa. Os questionários, portanto, possuíam questões retiradas da avaliação de alfabetização e questões elaboradas pelos pesquisadores. As perguntas (assim como os textos) foram previamente submetidas à professora da escola que auxiliou nas pesquisas para a sua aprovação. O material elaborado (texto e questionário) foi apresentado aos estudantes, que foram instruídos a ler e responder as questões colocadas. Não foi destinado tempo limite para completar a atividade, mas todos os alunos terminaram dentro do prazo máximo previsto.

Em uma segunda intervenção, foi repetida a proposta de leitura e interpretação de textos, porém dessa vez com o apoio da ferramenta Sobek. Os textos utilizados foram novamente retirados da *ANA*, sendo parte das perguntas elaboradas pelo pesquisador e parte original da avaliação. Tanto texto quanto questionário foram aprovados pela professora responsável previamente a realização da atividade. O minerador Sobek e o grafo produzido pelo minerador foram disponibilizados aos alunos para que fossem utilizados como bem entendessem.

Durante as atividades iniciais, não foi possível encontrar uma diferença estatisticamente significativa entre os resultados obtidos com o uso do minerador de textos Sobek e sem o uso do minerador. Esse resultado decorre tanto pelo baixo número de participantes da pesquisa (que encerrou com pouco mais de 10 alunos participando de todas as etapas) quanto pelos resultados similares obtidos pelos estudantes em ambas as intervenções. O baixo número de alunos presentes na última etapa, por sua vez, é justificado pelo final do calendário escolar que ocorreu muito próximo à última intervenção. Muitos estudantes estavam se preparando para provas finais ou tinham compromissos relacionados à suas formaturas de grau.

### 5.3.2.5 Avaliação dos estudantes

Após o término das atividades, foi apresentado um questionário aos alunos com o objetivo de verificar as percepções deles quanto ao uso e aplicação da ferramenta Sobek e sobre a atividade desenvolvida. O questionário foi escrito pelos pesquisadores para avaliar a percepção dos estudantes em relação a ambos os temas. As questões utilizaram respostas na escala Likert de sete ou cinco pontos com mensuração ordinal, na qual os rótulos da escala são exibidos de forma ordenada (Gil, 2009).

O questionário apresentado aos estudantes continha sete perguntas:

1. Qual o seu nível de satisfação em relação às atividades realizadas?
2. Qual o seu nível de satisfação em relação à ferramenta Sobek?
3. Você considera que o grafo representa o texto?
4. Qual o grau de dificuldade para utilizar a ferramenta Sobek?
5. Nas atividades de leitura, você considera que o Sobek ajudou a entender o texto?
6. Nas atividades de escrita, você considera que o Sobek ajudou a escrever o texto?
7. Você utilizaria o Sobek fora da sala de aula?

Para as duas primeiras questões, poderia ser associada uma avaliação na escala de 1 a 7, sendo o valor 1 equivalente à afirmação “*Achei muito ruim*” e o valor 7 equivalente à afirmação “*Gostei bastante*”. O questionário foi aplicado a todos os estudantes que estavam presentes na última atividade, totalizando dez respostas. Ressaltando a época em que a atividade final foi realizada, coincidindo com o final das atividades escolares, fez com que o número de estudantes presentes nessa atividade fosse menor que aquele que iniciou o experimento.

As três primeiras perguntas do questionário buscaram avaliar a percepção dos alunos quanto às atividades e a ferramenta Sobek de forma geral. A primeira questão do questionário identificou a motivação dos estudantes e a satisfação de modo geral com todas as atividades desenvolvidas.

Os resultados mostraram que 80% dos participantes consideraram-se muito ou bastante satisfeitos no que concerne às atividades propostas, enquanto 20% disseram estar satisfeito, como mostra a Figura 12. Seis dos dez estudantes deram classificação máxima para as atividades.

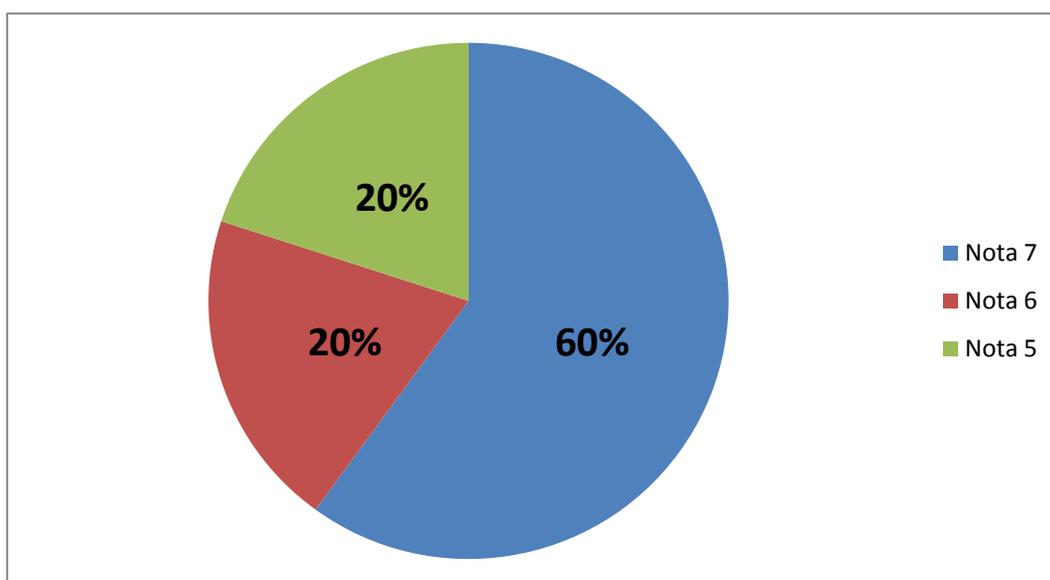


Figura 12 Gráfico referente ao índice de satisfação com o minerador Sobek e com as atividades desenvolvidas

A segunda pergunta buscou verificar se a ferramenta de mineração de textos Sobek agradou aos alunos. O resultado obtido foi o mesmo que para a pergunta de número 1. Todos manifestaram satisfação, sendo que 80% deles responderam ter gostado muito ou bastante do minerador de textos.

A terceira pergunta permitia uma resposta na escala de 1 a 5, sendo 1 equivalente à “*Representou muito mal*” e escala 5, “*Representou muito bem*”. Esta questão indagou sobre um aspecto considerado muito relevante para a pesquisa e obteve um alto índice de afirmação positiva. Ao serem questionados se o Sobek conseguiu representar bem o texto através de seu grafo, 70% dos estudantes respondeu que “*Representou muito bem*”, o que corrobora com a validação do aspecto de mineração de textos, descrito previamente. Nenhum aluno respondeu que o Sobek representou mal ou muito mal o texto. O gráfico com as respostas dos estudantes pode ser observado na Figura 13.

Por fim, a quarta questão permitia uma resposta na escala de 1 a 5, sendo 1 equivalente à “*Muito difícil*” e o valor 5, “*Muito fácil*”. Os resultados indicam que o Sobek é uma ferramenta de fácil utilização, tendo apenas um estudante considerado o Sobek difícil de utilizar.

Esses resultados são motivadores e demonstram que a percepção dos alunos foi positiva quanto às intervenções. Ademais, indicam que a ferramenta Sobek está

desenvolvida nos princípios pedagógicos propostos e que possui capacidade de ser utilizada amplamente por estudantes e professores. Pontos considerados fortes do Sobek, tais como a simplicidade de uso e a sua interface, foram reforçados pela avaliação recebida dos alunos.

As questões de 5 a 7 possuíam apenas 3 alternativas de respostas: sim, não, não sei. Estas questões visavam identificar se os usuários conseguiram utilizar o Sobek para auxiliá-los nas atividades indicadas. De forma indireta, estas questões também serviam para reforçar a avaliação do algoritmo proposto para o Sobek do ponto de vista de capacidade de extração de conceitos e representação de informação. Todos os alunos responderam que “*Sim*” quando perguntados se o Sobek os ajudou a compreender o texto. Isso implica que a forma como o Sobek apresenta os resultados, com termos e relacionamentos em destaque e na forma de um grafo, permite ao estudante observar questões que podem passar despercebidas no texto e pode facilitar a compreensão sobre aquilo que foi lido.

Nos demais questionamentos, apenas um estudante respondeu “*Não sei*”, tendo os demais respondido positivamente às perguntas. Isso demonstra que o Sobek foi capaz de identificar os termos mais relevantes do texto e que a representação de informação está organizada de forma a auxiliar os usuário. O minerador possibilitou a identificação e o relacionamento dos termos mais relevantes do texto, tendo a representação gráfica possibilitado aos alunos identificar tais informações extraídas pelo Sobek. Além disso, a interação com os termos do grafo permitiu aos estudantes organizar suas ideias e planejar a escrita de seus textos.

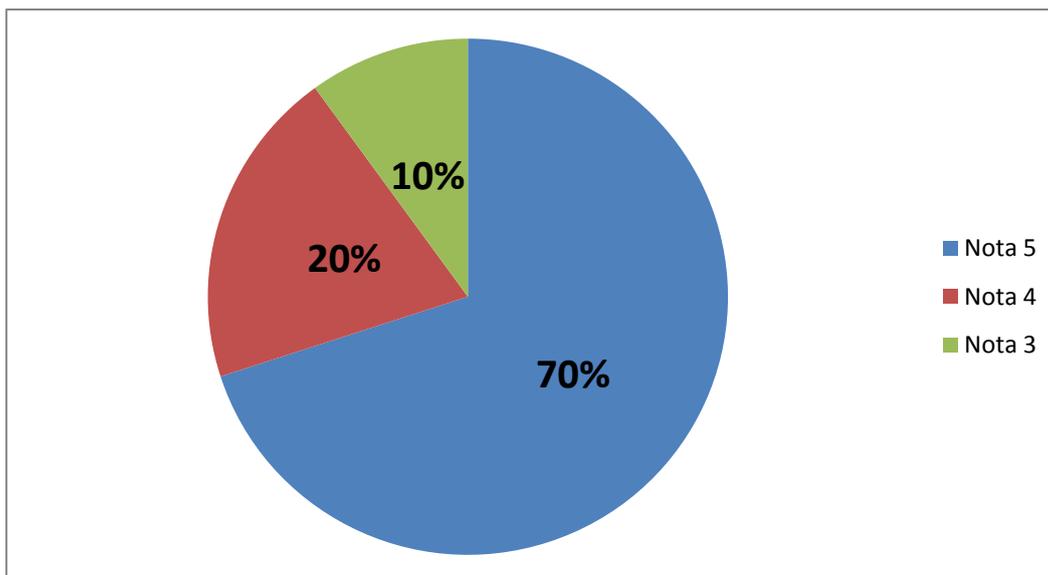


Figura 13 Nota dos alunos para a pergunta se o grafo do Sobek representa bem o texto

As respostas obtidas pelas questões de 5 a 7 demonstram que o uso do Sobek como minerador de texto e organizador gráfico tem implicações positivas junto aos estudantes e que o algoritmo utilizado por ele é eficiente tanto para auxiliar a compreensão textual quanto para apoiar o processo de produção textual. A representação gráfica desses termos também foi considerada adequada, enfatizando que auxiliou na compreensão textual dos materiais e ajudou os alunos a melhor organizar os textos que deveriam escrever. Isso demonstra que foi alcançado o propósito para o qual a interface gráfica do minerador foi planejada.

Por fim, a maioria dos estudantes afirmou que utilizaria o Sobek fora da sala de aula. Porém, em entrevista posterior com o pesquisador, eles não souberam afirmar como o Sobek poderia ser utilizado. Um dos motivos que dificulta o uso do minerador mais frequentemente é a falta de acesso à tecnologia por parte dos estudantes, que na sua maioria possuíam apenas dispositivos móveis como ferramentas tecnológicas.

Ao final do questionário, foram realizadas 3 perguntas de livre resposta aos estudantes. Essas perguntas permitiam ao aluno usar suas próprias palavras para descrever a atividade e ferramentas e expressar sua percepção quanto à pesquisa. Elas cobriam aspectos relacionados à ferramenta de modo geral e à atividade realizada. Não foi instruído um mínimo ou máximo de palavras a serem utilizadas e foi enfatizado que as respostas deveriam ser honestas, pois isso não afetaria o desempenho dos estudantes e ajudaria a aprimorar a ferramenta para que os demais alunos possam utilizar.

As perguntas do questionário de livre resposta foram:

8. Quais os pontos bons do Sobek?

9. Quais os pontos ruins do Sobek?

10. O que você achou das atividades?

De maneira geral, os alunos responderam de forma sucinta cada uma das perguntas. Para a pergunta 8, a grande maioria escreveu que o Sobek auxilia no aspecto de compreensão textual. O conceito mais comum utilizado pelos estudantes é que o Sobek “*ajuda a entender o texto*”. Todos os participantes expressaram satisfação quanto à capacidade do minerador de sintetizar o texto e apresentar informações e palavras relevantes para a compreensão do mesmo.

Considerando o âmbito deste trabalho, o retorno recebido dos estudantes foi muito positivo para avaliar o uso do Sobek como auxílio à compreensão textual.

Para a questão 9, diversos estudantes escreveram que não conseguiam apontar pontos negativos na ferramenta. Dos pontos negativos levantados, a maioria considerou o minerador difícil de utilizar no início, mas essa dificuldade desapareceu no decorrer das atividades. Dois alunos apontaram que o grafo gerado pelo Sobek não apresenta os detalhes da história, apenas o tema central e outro aluno comentou que considerou ruim a ‘*velocidade da ferramenta*’.

A atividade realizada foi elogiada por todos os estudantes. Diversos alunos destacaram que gostariam de continuar a realizar atividades como a ferramenta. Alguns consideraram que as intervenções os ajudaram a “*entender melhor os textos*” e que se “*divertiram bastante com o Sobek*”.

### **5.3.2.5 Avaliação da professora**

Após o encerramento das atividades, a professora contatada junto à escola que acompanhou as intervenções foi questionada para que realizasse uma avaliação de forma geral. Essa entrevista se deu através de troca de mensagens eletrônicas (*e-mail*) e de um questionário com perguntas de livre resposta. O questionário foi enviado à professora responsável para que preenchesse no momento que considerasse adequado.

Foram elaboradas 10 perguntas para a professora:

1. Qual a opinião do professor acerca da atividade realizada?
2. A ferramenta é simples de utilizar?
3. O professor usaria a ferramenta durante suas aulas?
4. Quais as principais dificuldades em preparar a atividade?
5. Qual a relação do professor com recursos tecnológicos pedagógicos?
6. Qual a avaliação do professor acerca do desempenho dos estudantes durante a atividade?
7. Toda a atividade transcorreu em sala de aula?
8. Algum aluno solicitou auxílio para o uso da ferramenta ou desenvolvimento da atividade?
9. Qual a avaliação geral do professor acerca dos aspectos pedagógicos da ferramenta?
10. Quais modificações poderiam ser realizadas na ferramenta para auxiliar o seu uso em sala de aula?

A primeira resposta da professora entrevistada apresenta sua aprovação quanto à atividade e a ferramenta. De acordo com a professora, *“a utilização do Sobek foi extremamente produtiva, principalmente com alunos que, até então, acreditavam ter mais dificuldade em produzir textos”*. A professora também comentou que os usos do computador e de ferramentas multimídia tornaram a atividade mais atrativa para os estudantes.

Foram destacadas as características dos textos utilizados nas atividades de compreensão textual, sendo eles descritos como textos mais simples e com menos relacionamentos entre os termos. A professora ressaltou que textos mais complexos (como aqueles apresentados na aula de história) podem apresentar grafos com mais relacionamentos. Neste caso, de acordo com a professora, o Sobek seria ainda mais útil para os alunos. Ainda de acordo com ela, o grafo apresentado pelo Sobek tem grande possibilidade de uso pedagógico, tanto para a produção quanto compreensão textual.

A professora também indicou nas suas respostas que considera que devem ser realizados mais encontros com os alunos para que estes consigam dominar a tecnologia apresentada. A principal dificuldade percebida foi a interação com o computador. Os estudantes tiveram dificuldades iniciais em escrever textos no editor de textos e em

manipular o grafo. Apesar disso, a professora mencionou que os estudantes obtiveram um bom desempenho nas atividades.

As respostas para as demais perguntas consideram que o Sobek é simples de utilizar e uma ferramenta interessante para utilização em sala de aula. Mesmo a professora não tendo muito contato com recursos tecnológicos, ela considerou o uso da ferramenta como sendo simples e fácil de usar.

As respostas ao questionário e a entrevista realizada com a professora evidenciou que sua percepção sobre o Sobek é de que a ferramenta tem bom potencial pedagógico e que as intervenções realizadas foram dosadas de forma aceitável para as turmas de alunos. A professora constantemente destacou em suas respostas o auxílio que a ferramenta trouxe para os alunos com relação a sua capacidade de se sentir apto e confiante, tanto na produção textual quanto na resposta de questionários de interpretação de texto. A capacidade do Sobek de encontrar termos relevantes em textos e a apresentação do grafo também foram enfaticamente elogiados.

O questionário completo realizado com a professora e suas respostas encontram-se no Apêndice D.

### **5.3.2.6 Conclusões sobre o estudo da fase preliminar**

De forma geral, os resultados obtidos foram bastante satisfatórios. Apesar de não ter sido encontrado diferença estatística no número de questões respondidas corretamente pelos estudantes quando apoiados pelo Sobek e quando respondendo da forma convencional, a ferramenta se mostrou importante do ponto de vista pedagógico. A professora que acompanhou as intervenções e os alunos que participaram dela consideraram o minerador como uma ferramenta capaz de auxiliar no processo de letramento e aprovaram o seu uso em sala de aula. A grande aprovação por parte dos estudantes permite considerar que o uso do Sobek também é um incentivo para que eles se engajem de forma mais ativa durante as aulas e que o foco de aprendizagem seja no indivíduo que está manipulando o grafo.

Considerando as respostas da professora, vemos que há um potencial para o uso em sala de aula da ferramenta. Porém, uma das principais dificuldades encontradas nesse estudo foi justamente o uso do computador em sala de aula pelos estudantes.

Devido à natureza das atividades, era essencial que os alunos já tivessem tido contato com um computador e fossem capazes de manipulá-lo de forma adequada. Inicialmente, considerando a destreza dos jovens de hoje com a tecnologia, foi considerado que todos os alunos já tivessem utilizado alguma vez o computador. Logo no primeiro encontro essa premissa se mostrou inválida, sendo bastante complicado para muitos alunos executarem as tarefas requeridas. As dificuldades ocorreram tanto no manuseio do mouse e interação com o grafo quanto na escrita de textos com o teclado. Para o uso da ferramenta em sala de aula, seria necessário que a turma de estudantes soubesse utilizar tais meios de interação com o computador, coisa que não ocorre na escola onde a atividade foi realizada. Essa dificuldade foi rapidamente diagnosticada no primeiro encontro e um auxílio na questão tecnológica foi fornecida aos estudantes. Apesar da dificuldade em utilizar o computador, vale destacar a facilidade e rapidez com que os mesmos manipulavam dispositivos móveis, indicando que essa deve ser a continuação lógica do desenvolvimento da ferramenta Sobek como forma de apoio ao letramento.

Mesmo em vista das dificuldades, os estudantes aprovaram a ferramenta e atividades. Muitos expressaram que gostariam de continuar as atividades mesmo após o término do ano letivo e por várias vezes foi indagado se era possível utilizar a ferramenta em casa. A professora apoiou em todos os aspectos as intervenções e propostas do pesquisador e demonstrou interesse em continuar a pesquisa com turmas no próximo ano letivo. Aqui vale enfatizar que a escola de maneira geral e a professora responsável foram de auxílio inestimável para a realização das intervenções.

### **5.3.3 Estudo da fase final**

O estudo central deste trabalho foi realizado utilizando a ferramenta de mineração de textos Sobek como forma de apoio à compreensão textual. Tendo este enfoque, a pesquisa utilizou uma coletânea de textos e perguntas referentes aos textos para avaliar como o uso do Sobek interferiu na capacidade do estudante em escolher a alternativa que corretamente correspondia à informação do texto. O estudo é composto de uma análise quantitativa e qualitativa, a exemplo daquele realizada no estudo preliminar.

#### **5.3.2.1 Sujeitos da pesquisa**

A pesquisa foi realizada no Colégio Murialdo, em Caxias do Sul. Os estudantes participantes da pesquisa integram um projeto social chamado “*Ação Social*”, do qual

participam crianças e adolescentes de famílias de baixa renda. Os integrantes do projeto *Ação Social* têm reforço acadêmico e ensino profissionalizante no turno em que não possuem aula, permanecendo grande parte do dia na escola. Participaram da pesquisa 43 alunos, 23 do sexo masculino e 20 do sexo feminino.

Os alunos que participaram do estudo estão inseridos no primeiro e terceiro ciclo do projeto “*Ação Social*” e correspondem a estudantes de quinto ano e oitavo ano. Todos os estudantes estavam familiarizados com o uso de computadores de mesa e participaram de forma voluntária da pesquisa. A todos foi entregue um termo de consentimento e foi explicado o propósito da pesquisa e da ferramenta de mineração de textos Sobek.

Assim como no estudo preliminar, os estudantes assinaram o termo de consentimento para realizar a atividade, e o diretor da Escola assinou um documento permitindo a realização da pesquisa dentro da escola. Estes termos encontram-se nos Apêndices *E* e *F*, respectivamente.

### **5.3.2.2 Materiais das intervenções**

As intervenções foram realizadas utilizando textos retirados de apostilas destinadas aos anos correspondentes a cada ciclo e textos retirados da Avaliação Nacional de Alfabetização. Cada texto era composto de perguntas retiradas do mesmo material. Além dos textos utilizados para avaliação da compreensão textual dos estudantes, um questionário semelhante ao apresentado no estudo preliminar foi entregue a eles no final das atividades. Todos os textos e questionários encontram-se no Apêndice G.

O minerador de textos Sobek utilizado nas intervenções sofreu alterações em relação ao utilizado nas atividades na Escola Municipal Lidovino Fanton. Com base no estudo preliminar, foi possível identificar pontos fortes e fracos da ferramenta. Um dos principais pontos levantado pelos estudantes está relacionado à velocidade de mineração e interação com o grafo. Foi apontado que, por vezes, a ferramenta pode tornar-se lenta ou demorar em realizar a extração de termos. Como forma de solucionar tal problema, foi realizada uma alteração no código fonte do Sobek, removendo funções desnecessárias e tentando reduzir a necessidade de processamento gráfico ao exibir o grafo resultante da mineração. Isso permitiu que fosse possível utilizar o Sobek em

computadores menos avançados, reduzindo a percepção do estudante do atraso na resposta da ferramenta.

Outra atualização realizada foi no intuito de coletar uma maior quantidade de dados acerca da interação do estudante com o grafo. Foi adicionada ao Sobek uma função que salva todas as interações que são realizadas sobre o grafo, seja uma adição, remoção ou edição de nodo ou aresta. Além disso, a organização que o usuário faz ao movimentar os nodos também é armazenada. Com isso, é possível analisar como os estudantes interagem com cada nodo apresentado pelo Sobek, quais termos são desnecessários para a compreensão do texto (e, portanto, removidos do grafo) e quais termos o Sobek deveria apresentar no grafo e não apresentou (sendo estes os termos adicionados ao grafo pelo usuário).

Ao final da atividade, Sobek envia no formato *JSON* a lista de ações dos usuários para o servidor do grupo de pesquisa, que se encontra na UFRGS. Cada arquivo é identificado por um número único, com data e hora de início, de fim e de cada interação com o grafo. Com base neste arquivo, é possível inferir como o estudante se portou frente ao grafo, o que foi realizado primeiro por cada aluno e quais ações foram mais executadas durante as atividades. Também é possível analisar como estudantes que obtiveram um bom (ou mal) desempenho respondendo as questões do texto interagiram com o grafo do Sobek.

### **5.3.2.3 Descrição do experimento**

A primeira de atividade foi desenvolvida utilizando dois textos para avaliar a capacidade dos estudantes em responder ao questionário com perguntas de interpretação de texto. Esta atividade foi realizada instruindo os estudantes a lerem os textos e responderem as perguntas. Não foi especificado tempo mínimo ou máximo para a realização da mesma. Nesta primeira etapa não havia o apoio de ferramentas externas para auxiliar na compreensão textual, sendo apenas fornecido o texto como instrumento para a resposta do questionário.

A atividade inicial foi realizada com ambos os ciclos, cada qual com seu conjunto de textos e perguntas. A segunda intervenção foi realizada na sala de informática do Colégio Murialdo. Primeiramente, esta intervenção ocorreu junto aos estudantes de primeiro ciclo (quinto ano). Ela foi efetivada durante o período de recesso dos alunos para almoço, entre as aulas do período da manhã e o início do projeto social

à tarde. A sala de aula utilizada para atividade conta com 7 computadores. Dessa forma, os alunos foram chamados de 7 em 7, tendo um computador para cada aluno.

Para cada turma de estudantes, foi realizada uma apresentação sobre o software Sobek e uma demonstração sobre como utilizá-lo. Foi também explicada a atividade, pesquisa e objetivos daquela intervenção. Os estudantes já haviam recebido durante a primeira atividade o termo de consentimento e estavam cientes que a atividade era facultativa e não iria interferir no restante das atividades escolares. Cada estudante sentou à frente de um computador e realizou a tarefa de forma individual. A professora Simone Emer (responsável pela diretoria pedagógica do colégio) acompanhou e auxiliou durante todas as atividades. Ela também prestou auxílio aos estudantes quando surgiram dúvidas e os organizou para que pudessem realizar as atividades no momento considerado mais apropriado.

A atividade consistiu em acessar um formulário armazenado no *Google Forms*, que continha as instruções, textos e questionamentos, que deveriam ser lidos e respondidos pelos alunos. Para cada um dos dois textos a serem lidos, havia uma sequência de perguntas que deveriam ser respondidas. Só era possível prosseguir do primeiro para o segundo texto após ter respondido todas as questões.

Em paralelo ao navegador web que apresentava os textos, os alunos foram instruídos a utilizar o Sobek para extrair os termos principais do texto e manipular o grafo. Foi requisitado que a atividade fosse realizada na seguinte ordem:

1. Leia o texto na íntegra
2. Copie e cole o texto no minerador de textos Sobek
3. Extraia o grafo do texto
4. Adicione um termo que você considera relevante para a compreensão do texto e que não está no grafo
5. Remova um termo que está no grafo e que você considera que não é relevante para a compreensão do texto
6. Edite algum relacionamento entre termo
7. Realize quaisquer alterações no grafo que você considera necessário para que o grafo represente o texto
8. Responda as perguntas do questionário

Após os estudantes de primeiro ciclo terem concluído as atividades, os estudantes de terceiro ciclo (oitavo ano) foram chamados a participar do experimento. Neste segundo caso, a sala de informática utilizada possuía 20 computadores, sendo realizada apenas uma rodada de atividades. Cada estudante sentou em um computador individual e o procedimento descrito acima foi repetido. Novamente, os textos e questionamentos foram retirados de apostilas destinados ao oitavo ano e da *ANA*. Após o término da atividade, foi requisitado que os estudantes respondessem às perguntas sobre suas percepções acerca da atividade e do minerador de textos Sobek.

Os textos e questionários utilizados junto aos estudantes do Ciclo 1 encontram-se no Apêndice H e aqueles utilizados junto aos alunos do Ciclo 3 encontram-se no Apêndice I. O questionário acerca da atividade e do minerador de textos Sobek encontra-se no Apêndice J.

## **6. Resultados e discussão**

Nesta sessão estão descritos os resultados obtidos a partir das intervenções realizadas no Colégio Murialdo. Os dados coletados são analisados por diversos prismas, buscando não apenas identificar se o uso do minerador de textos Sobek contribuiu para a compreensão textual dos estudantes mas também como a interação com o minerador ocorreu. O questionário aplicado aos estudantes no final da intervenção indaga sobre suas percepções acerca do Sobek e da atividade realizada.

### **6.1 Análise dos dados coletados referentes ao estudo acerca da influência do minerador de textos Sobek na compreensão textual dos estudantes**

O estudo conduzido teve por objetivo avaliar se a utilização do minerador de textos Sobek auxilia os estudantes na compreensão textual. Nesse aspecto, foi comparado o número de questões de interpretação de texto respondidas corretamente pelos estudantes quando não havia o suporte da ferramenta com a situação em que a ferramenta Sobek foi empregada para a realização da atividade.

O primeiro resultado apresentado corresponde às intervenções realizadas com alunos do primeiro ciclo (quinto ano). A pesquisa foi realizada com um total de 24 alunos presentes na primeira intervenção, na qual o Sobek não foi utilizado, e 27 alunos presentes na segunda intervenção, onde o questionário foi respondido com auxílio da ferramenta. Para a avaliação na qual não foi utilizado o minerador de textos Sobek, foram disponibilizados dois textos para os estudantes. O primeiro continha duas questões objetivas, com 4 alternativas cada; o segundo texto, 3 perguntas objetivas e 4 alternativas cada. Foi instruído aos estudantes que deveriam marcar apenas a alternativa correta. Não foi estipulado um tempo máximo ou mínimo no qual a atividade deveria ocorrer, deixando a cargo de cada estudante entregar o questionário quando lhe fosse conveniente.

O resultado com o número de acertos das questões do primeiro texto está apresentado no quadro 5.

**Quadro 5: Situação de controle para ciclo 1: número de alunos que responderam corretamente o primeiro texto sem apoio do Sobek**

<b>Situação de Controle: Avaliação a partir do primeiro texto sem apoio do Sobek</b>		
	Número de alunos que respondeu corretamente	Percentual de acertos
Questão 1	6 de 24 alunos	25%
Questão 2	16 de 24 alunos	66%

É possível observar um grande número de respostas incorretas para as questões, sobretudo considerando-se que tratava-se de textos retirados da ANA. A segunda questão corrobora com os resultados obtidos para o primeiro texto. O quadro 6 apresenta os resultados e médias de respostas corretas para cada questão.

**Quadro 6: Situação de controle para ciclo 1: número de alunos que responderam corretamente o segundo texto sem apoio do Sobek**

<b>Situação de Controle: Avaliação a partir do segundo texto sem apoio do Sobek</b>		
	Número de alunos que respondeu corretamente	Percentual de acertos
Questão 1	4 de 24 alunos	16%
Questão 2	20 de 24 alunos	83%
Questão 3	12 de 24 alunos	50%

Considerando a média de respostas corretas em cada texto, 45% das questões do texto 1 foram respondidas corretamente e 50% das questões do texto 2 foram respondidas corretamente. Tais resultados evidenciam a dificuldade dos estudantes em interpretar textos, problema que pode ser minimizado por uma ferramenta de apoio à compreensão textual.

A intervenção seguinte realizada com a turma de alunos do quinto ano utilizou o Sobek para apoiar os estudantes na compreensão dos textos. Nesta intervenção, os alunos tiveram o apoio do minerador e deveriam interagir com a ferramenta antes de responder as perguntas, conforme detalhado na sessão 5.3.2.3. Os resultados obtidos em cada um dos dois textos são apresentados nos quadros 7 e 8.

**Quadro 7: Situação experimental para ciclo 1: número de alunos que responderam corretamente o primeiro texto com apoio do Sobek**

<b>Situação Experimental: Avaliação a partir do primeiro texto com apoio do Sobek</b>		
	Número de alunos que respondeu corretamente	Percentual de acertos
Questão 1	22 de 27 alunos	81%
Questão 2	13 de 27 alunos	48%
Questão 3	18 de 27 alunos	66%

**Quadro 8: Situação experimental para ciclo 1: número de alunos que responderam corretamente o segundo texto com apoio do Sobek**

<b>Situação Experimental: Avaliação a partir do segundo texto com apoio do Sobek</b>		
	Número de alunos que respondeu corretamente	Percentual de acertos
Questão 1	19 de 27 alunos	70%
Questão 2	16 de 27 alunos	59%
Questão 3	13 de 27 alunos	48%

É possível observar que a média de questões respondidas corretamente com o auxílio do minerador de textos Sobek é mais alta que sem o minerador. Para o primeiro texto, os estudantes responderam corretamente em média 65% das questões corretamente; no segundo texto, a média de respostas corretas corresponde a 59% das questões.

Uma série de intervenções semelhantes à aplicada ao ciclo 1 foi realizada com alunos do ciclo 3 (oitavo ano). A turma de estudantes presente contou com 16 alunos para a atividade de pesquisa inicial (onde o Sobek não foi utilizado) e 12 alunos na atividade de pesquisa com o minerador Sobek. De forma análoga à atividade

apresentada anteriormente, primeiro foi realizada uma intervenção na qual os estudantes deveriam responder às perguntas sem o apoio da ferramenta de mineração Sobek. Os resultados obtidos pelos alunos estão apresentados no quadro 9. O segundo texto da atividade possuía questões formuladas pelos pesquisadores e utilizadas no estudo preliminar, além daquelas já existentes na ANA.

**Quadro 9: Situação de controle para ciclo 3: número de alunos que responderam corretamente o primeiro texto sem apoio do Sobek**

<b>Situação de Controle: Avaliação a partir do primeiro texto sem apoio do Sobek</b>		
	Número de alunos que respondeu corretamente	Percentual de acertos
Questão 1	6 de 16 alunos	37%
Questão 2	6 de 16 alunos	37%

**Quadro 10: Situação de controle para ciclo 3: número de alunos que responderam corretamente o segundo texto sem apoio do Sobek**

<b>Situação de Controle: Avaliação a partir do segundo texto sem apoio do Sobek</b>		
	Número de alunos que respondeu corretamente	Percentual de acertos
Questão 1	9 de 16 alunos	56%
Questão 2	9 de 16 alunos	56%
Questão 3	7 de 16 alunos	43%
Questão 4	8 de 16 alunos	50%
Questão 5	8 de 16 alunos	50%

A partir da avaliação do número médio de respostas corretas para cada texto, os estudantes acertaram 37% das questões do primeiro texto e 51% das questões do segundo texto. Apesar de acertarem acima da metade de questões do segundo texto, estes resultados ainda evidenciam uma certa dificuldade dos estudantes em questões de compreensão textual.

A segunda intervenção realizada com os estudantes do Ciclo 3 ocorreu com o uso do minerador de textos Sobek como forma de apoio à compreensão textual. Nesta

ocasião, foi realizada a atividade com 12 alunos e o número de respostas corretas para cada texto está apresentada no quadro 11 e 12.

**Quadro 11: Situação experimental para ciclo 3: número de alunos que responderam corretamente o primeiro texto com apoio do Sobek**

<b>Situação Experimental:</b> Avaliação a partir do primeiro texto com apoio do Sobek		
	Número de alunos que respondeu corretamente	Percentual de acertos
Questão 1	12 de 12 alunos	100%
Questão 2	7 de 12 alunos	58%
Questão 3	5 de 12 alunos	41%

**Quadro 12: Situação experimental para ciclo 3: número de alunos que responderam corretamente o segundo texto com apoio do Sobek**

<b>Situação Experimental:</b> Avaliação a partir do segundo texto com apoio do Sobek		
	Número de alunos que respondeu corretamente	Percentual de acertos
Questão 1	8 de 12 alunos	66%
Questão 2	5 de 12 alunos	41%
Questão 3	6 de 12 alunos	50%
Questão 4	12 de 12 alunos	100%

É possível observar que o número médio de respostas corretas quando utilizando o Sobek subiu para 66% no primeiro texto e 64% no segundo texto.

Para confirmar que há uma diferença estatisticamente significativa na quantidade de respostas corretas na intervenção sem o uso do Sobek e naquela onde o Sobek é utilizado, foi realizada a análise de variância (ANOVA), utilizando funções específicas para o cálculo disponíveis no pacote Microsoft Excel. Os dados que serviram de entrada para a análise foram a quantidade de respostas corretas que cada aluno forneceu em cada intervenção. Ou seja, considerando um estudante aleatório  $A$ , os dados são a porcentagem de questões respondidas corretamente por  $A$  na primeira intervenção e na segunda.

A análise estatística necessária para avaliar se há relação entre o uso do Sobek e a porcentagem de acertos dos estudantes exige duas informações: a correlação entre os dados e o *valor de p*. A correlação (ou *correlação de Pearson*) é uma forma de afirmar se há uma relação estatística entre duas variáveis. O coeficiente de correlação pode ter valores de -1 até 1, onde:

- O valor “0” significa que não há relação entre as variáveis;
- Valores negativos indicam relação negativa entre as variáveis;
- Valores positivos indicam que um valor está positivamente relacionado ao outro.

O *valor de p* (do inglês, *p-value*), conhecido como significância estatística, indica a chance ou a probabilidade da diferença observada entre as categorias (no caso deste estudo, da quantidade de acertos com e sem o uso do minerador Sobek) ser devido ao acaso, e não aos fatores que estão sendo estudados. Adotou-se nesta pesquisa o critério que estabelece como significativo um *valor de p* menor que “0,05”, ou seja, assumiu-se uma margem de erro de 5%.

Analisando estatisticamente os resultados obtidos por cada estudante nas situações de controle e experimental, foi calculado o coeficiente de correlação entre o uso do Sobek e a porcentagem de respostas corretas pelos alunos. Todos os dados coletados foram levados em consideração, ou seja, o percentual de acertos para ambos os ciclos na primeira e na segunda intervenção. Essa métrica corresponde ao coeficiente de correlação e *valor de p* da pesquisa, uma vez que o uso do Sobek para compreensão textual é advogado como sendo positivo para a compreensão textual.

O coeficiente de correlação encontrado é de “0,84”, ou seja, há uma forte correlação entre o uso do Sobek e o percentual de respostas corretas dos estudantes. Isso é um resultado que corrobora com a proposta deste trabalho, pois demonstra que a utilização do Sobek teve influência no número de respostas corretas fornecidas pelos alunos. O *valor de p* deste teste é de “0,0085”. Este valor está bem abaixo daquele considerado limite superior para aceitar que há uma significância estatística nos dados. Assim, podemos concluir que o minerador de textos Sobek de fato influenciou positivamente no percentual de respostas corretas fornecidas pelos estudantes durante a intervenção.

Analisando individualmente as intervenções para cada ciclo, podemos medir a correlação entre o uso do minerador Sobek e o número de respostas corretas que os alunos responderam. Para os estudantes de primeiro ciclo, foi possível apontar que a correlação entre as variáveis aqui expostas é de  $0,48$ . Isso indica que há uma correlação positiva entre o uso do Sobek e a porcentagem de acertos dos estudantes. A análise também aponta que o *valor de p* neste estudo é de  $0.000347598$ , ou seja, um valor bem menor que o valor mínimo esperado para que o resultado fosse considerado estatisticamente significativo. Com base nesse resultado, é possível afirmar que o Sobek teve um impacto positivo na quantidade de questões respondidas corretamente pelos estudantes.

Para os estudantes do ciclo 3, a análise estatística mostra que a correlação entre porcentagem de respostas corretas e o uso do Sobek é de “ $0,36$ ”. O valor também indica correlação positiva em relação ao número de respostas corretas dos estudantes. Porém, para estas intervenções, o *valor de p* encontrado é de “ $0,087$ ”, valor acima do valor mínimo estipulado para a margem de erro. Por tratar-se de estudantes de um nível mais avançado (ciclo 3), percebe-se que a ferramenta pode oferecer algum apoio nas questões de compreensão textual, porém não de forma tão evidente como no caso dos estudantes do primeiro ciclo. Estes alunos já possuem técnicas de leitura e compreensão textual mais bem desenvolvidas, reduzindo o impacto que o uso do minerador de textos Sobek têm no auxílio à compreensão textual. O resultado reforça aqueles encontrados por Manoli e Papadopoulou (2012) que afirmam que estudantes com baixa capacidade de leitura se beneficiam mais do uso de organizadores gráficos. Apesar de reduzido, é possível afirmar que o uso do Sobek tem influência positiva sobre o número de acertos dos estudantes com alta probabilidade de assertividade.

## **6.2 Análise da ação dos usuários sobre o grafo**

Após o estudo preliminar, foi verificada a necessidade de se obter mais dados referentes às ações dos usuários durante a interação com o minerador Sobek. Uma vez que o objetivo da pesquisa está relacionado à premissa que o minerador de textos Sobek auxilia os estudantes na compreensão textual, é importante identificar como o estudante interage com o grafo e que ações são mais frequentemente realizadas. Através desta análise, é possível apontar as ações mais comuns e menos comuns dos usuários e

identificar os pontos fortes e fracos do grafo no apoio às atividades de compreensão textual.

Uma vez que seria impossível para o pesquisador acompanhar todas as interações realizadas sobre o grafo para cada aluno durante a intervenção, se tornou necessário desenvolver uma ferramenta tecnológica para isso. Conforme descrito na sessão 5.3.2.4, foi desenvolvida uma função dentro do minerador Sobek que armazena todas as atividades do usuário após o grafo ter sido gerado. As ações sobre o grafo previstas são:

- Adição de nodo
- Remoção de nodo
- Adição de aresta
- Remoção de aresta
- Seleção de nodo

As ações de adição e remoção são relevantes para identificar como o usuário interage com os termos (se ele primeiro adiciona todos os termos e depois remover aqueles que não necessitam, por exemplo) e qual a forma como é pensada a edição do grafo para torná-lo tão representativo do texto quanto possível. Porém, é importante destacar também a ação de seleção de nodo. Ao clicar sobre um termo para selecioná-lo, são destacados para o usuário os vizinhos diretos desse nodo (aqueles nodos com o qual há uma aresta de ligação), bem como os trechos do texto onde o termo contido no nodo aparece. Isso fornece subsídios para que seja possível compreender melhor um termo, observando o seu uso no texto e os demais termos que fazem ligação com ele. Importante destacar que para reordenar os nodos no grafo não é necessário selecioná-los. Basta arrastar um nodo para a posição desejada. Assim, a captura da ação de selecionar um nodo apenas ocorre quando o usuário de fato o seleciona um e não quando ele passa o mouse por cima ou quando o arrasta.

Cada vez que o usuário executa alguma das ações mencionadas, o minerador armazena a ação, seu resultado e a ordem com que ela foi executada (a primeira ação tem ordem 1, a segunda ordem 2, e assim por diante). Ao final da atividade, quando o estudante encerra o grafo ou o aplicativo, essas ações são enviadas para serem armazenadas em um servidor, utilizando formato JSON. Um exemplo de arquivo contendo essas informações é apresentado nas figuras 14 e 15.

```

{"random":434630,"addNodeActions":[{"order":1,"nodeName":"anormal"}, {"order":2,"nodeName":"identidade"}, {"order":3,"nodeName":"condominio"}
,{"order":4,"nodeName":"homem"}], "removeNodeActions":[{"order":12,"nodeName":"frankestein"}, {"order":13,"nodeName":"frankestein"}], "addEdge
Actions":[{"order":6,"originNodeName":"homem", "destinyNodeName":"anormal"}, {"order":14,"originNodeName":"Frankenstein", "destinyNodeName"
:"condominio"}], "removeEdgeActions":[{"order":4,"originNodeName":"identidade", "destinyNodeName":"anormal"}, {"order":5,"originNodeName"
:"frankestein", "destinyNodeName":"homem"}], "nodeClickActions":[{"order":9,"nodeName":"frankestein"}, {"order":10,"nodeName":"frankestein"}
,{"order":11,"nodeName":"frankestein"}]}

```

Figura 14 Imagem contendo o JSON produzido pelo Sobek a partir da interação de um estudante com o grafo

```

1 { "random": 434630,
2   "addNodeActions":
3   [
4     {"order": 1, "nodeName": "anormal"},
5     {"order": 2, "nodeName": "identidade"},
6     {"order": 3, "nodeName": "condominio"}
7   ],
8   "removeNodeActions": [
9     {"order": 12, "nodeName": "frankestein"},
10    {"order": 13, "nodeName": "frankestein"}
11  ],
12  "addEdgeActions": [
13    {"order": 6, "originNodeName": "homem", "destinyNodeName": "anormal"},
14    {"order": 14, "originNodeName": "Frankenstein", "destinyNodeName": "condominio"}
15  ],
16  "removeEdgeActions": [
17    {"order": 4, "originNodeName": "identidade", "destinyNodeName": "anormal"},
18    {"order": 5, "originNodeName": "frankestein", "destinyNodeName": "homem"}
19  ],
20  "nodeClickActions": [
21    {"order": 9, "nodeName": "frankestein"},
22    {"order": 10, "nodeName": "frankestein"},
23    {"order": 11, "nodeName": "frankestein"}
24  ]
}

```

Figura 15 Representação visual da Figura 14

Na figura 15, podemos observar que as primeiras ações do usuário foram de adicionar termos que não estavam no grafo e que o usuário considerou relevante para a compreensão do texto. Para que o usuário possa identificar a necessidade desses nodos, ele deve primeiro ler todas as informações apresentadas no grafo, o que o auxiliaria a refletir sobre o texto. O mesmo ocorre para a ação de remoção de nodo. Uma vez que o usuário identifique que um termo não faz parte do conjunto de palavras necessárias para corretamente representar o texto, ele remove esse termo. Para isso, é necessário refletir sobre o texto e pensar do ponto de vista do grafo se o termo deve ou não estar presente.

Após adicionar os nodos, o usuário passa a editar as arestas, removendo aquelas que considera inadequadas. As ações realizadas sobre as arestas (ligações entre termos) necessita uma reflexão não apenas sobre o texto, mas também sobre os termos de forma

individual. Os relacionamentos podem representar eventos presentes no texto, mas também ligações que ocorrem entre termos (por exemplo, ligações de termos pelo seu significado semântico). Essa reflexão para alterar os relacionamentos presentes no grafo é fundamental para a compreensão textual, já que exige que as informações sobre os termos estejam bem consolidadas e que haja uma reflexão sobre um conjunto de ideias presentes no texto, e não apenas sobre termos isolados.

Utilizando os dados coletados pela ferramenta, foi analisado como os estudantes de cada ciclo interagiram sobre o grafo de maneira geral. Os estudantes do Ciclo 1 utilizaram predominantemente a funcionalidade de adição de nodo ao grafo. A ação de adicionar um nodo é 5 vezes mais utilizada que a ação de remoção de nodo e 6 vezes mais recorrente que ações de edição de relacionamentos. Além disso, essa ação de incluir um termo ao grafo é a primeira e segunda ação tomada pelos estudantes na grande maioria das vezes (aproximadamente de 85% das vezes).

É possível interpretar esse resultado como uma insegurança do estudante em contrariar as informações providas pela ferramenta. O aluno consegue percorrer o grafo e identificar termos que aparecem no texto e não aparecem no grafo, tornando o processo de adição de nodo semelhante a uma transferência de informação do texto para o grafo. Porém, no processo de remover um nodo, é necessário que o aluno confronte a informação fornecida pelo Sobek e que ele tenha certeza que aquele termo não deveria estar presente. É necessário não apenas uma reflexão sobre o texto, mas a capacidade do aluno de tomar a decisão de afirmar que algumas das informações contidas no grafo não correspondem a sua própria forma de compreensão/interpretação do texto. Essa ação demanda confiança do estudante sobre o que ele está lendo e sobre a ferramenta com que está interagindo, o que nem sempre ocorre, sobretudo com leitores menos experientes. Assim, o caminho mais seguro para o aluno é adicionar as informações que ele percebe que deveriam estar no grafo, sem alterar aquelas já presentes.

A ação de adicionar um nodo ao grafo demonstra que o estudante foi capaz de refletir sobre o texto e relacionar a informação presente no seu corpo de conhecimento com aquela apresentada pelo Sobek. A adição de termos e relacionamentos entre eles representa o conhecimento do aluno sendo transferido para o grafo. Tal ação mostra que o estudante foi capaz de identificar que há um relacionamento entre termos ou que há informações relevantes ao contexto do texto que não estavam presentes. Esse resultado

corroborar com aqueles identificados nas pesquisas utilizando mapas conceituais (Moreira, 2010; Ausubel, 2002), nos quais os mapas são utilizados para auxiliar na aprendizagem de conceitos através do relacionamento dos mesmos com outros já presentes na estrutura cognitiva do aprendiz. A associação de ideias e conceitos é representada de forma explícita através de arestas no grafo, representando também o conhecimento do aluno e dando subsídios para que ele encontre subsunções adequadas para ancorar as novas informações presentes no texto.

Além das informações obtidas acerca da manipulação do grafo pelos estudantes, durante as atividades foi solicitado aos participantes que salvassem os grafos produzidos por eles após responderem as questões do texto. Uma vez que o armazenamento da imagem do grafo foi realizado após o término do questionário de cada texto, essa ação não interferiu nas respostas dos estudantes. O próprio minerador Sobek possui um botão na interface de grafo que permite salvar o grafo como imagem de maneira simples. Com isso, é possível visualizar como estavam distribuídos os termos no grafo de cada aluno após a edição e não apenas quais termos e relacionamentos existiam (o que é possível obter através do arquivo JSON com a relação das ações do usuário sobre o grafo).

O grafo apresentado automaticamente pelo Sobek posiciona os termos em posições aleatórias, por vezes com as conexões embaralhadas e difíceis de visualizar. Parte da ação de movimentar os nodos permite que o usuário interaja com cada termo extraído, mesmo que seja apenas pelo breve momento de realocar ele em alguma posição da tela. Além do próprio termo, ao passar o mouse sobre um nodo ou arrastar ele, os seus vizinhos diretos são destacados, auxiliando na reflexão sobre o termo em relação aos demais no texto e facilitando o reposicionamento deles de forma que o grafo fique o mais organizado possível.

O Sobek busca extrair todos os termos que sejam relevantes para a compreensão do texto e que possam ser encontrados estatisticamente. Tendo em vista a capacidade da ferramenta em realizar esta seleção de termos, nem sempre é trivial para o aluno remover ou adicionar nodos ao grafo, uma vez que grande parte dos termos relevantes já devem estar elencados pelo Sobek. Conforme mencionado anteriormente, adicionar e remover nodos demanda uma reflexão sobre o texto antes de interagir com o grafo. Por vezes, o usuário tem dificuldades nessa ação. Um estudante do primeiro ciclo comentou este fato: *“todos as palavras que eu pensei já estão aqui”*. Mesmo nestes casos, é

importante fazer com que o estudante reflita sobre os termos do grafo e as relações nele apresentadas.

Para a intervenção realizada junto ao ciclo 1, foram escolhidos textos cuja representação gráfica pelo Sobek eram de simples visualização, com termos bem conectados e muitos nodos. A Figura 16 ilustra o resultado da mineração do texto 1 que os alunos deveriam ler para responder o questionário. Este primeiro texto chamado de “*O dono da bola*”, de autoria da escritora Ruth Rocha, versa sobre um jogador (chamado de Carlos Alberto e apelidado de “Caloca”) que sempre buscava levar vantagem no jogo por ser o dono da bola.

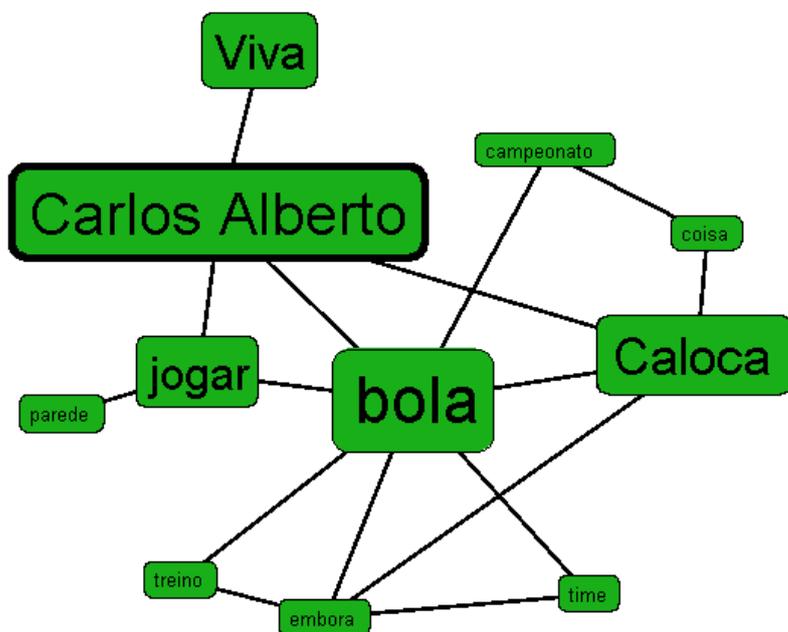


Figura 16 Grafo formado pelo Sobek a partir do texto "O dono da bola" de Ruth Rocha

Na Figura 17, o grafo exposto pelo Sobek foi reorganizado e teve termos adicionados e removidos pelo estudante A. Este estudante acertou todas as respostas para as questões da atividade e por isso seu grafo foi escolhido para análise. Apesar de haver poucos termos adicionados e removidos em relação ao grafo original, é possível observar que a organização do mesmo está diferente daquele grafo produzido pelo Sobek. Isso demonstra que houve uma interação do usuário com o grafo, evitando que termos se sobrepusessem. Os termos novos foram adicionados após a ordenação do grafo, uma vez que estão dispostos sem relacionamento e acima de arestas.

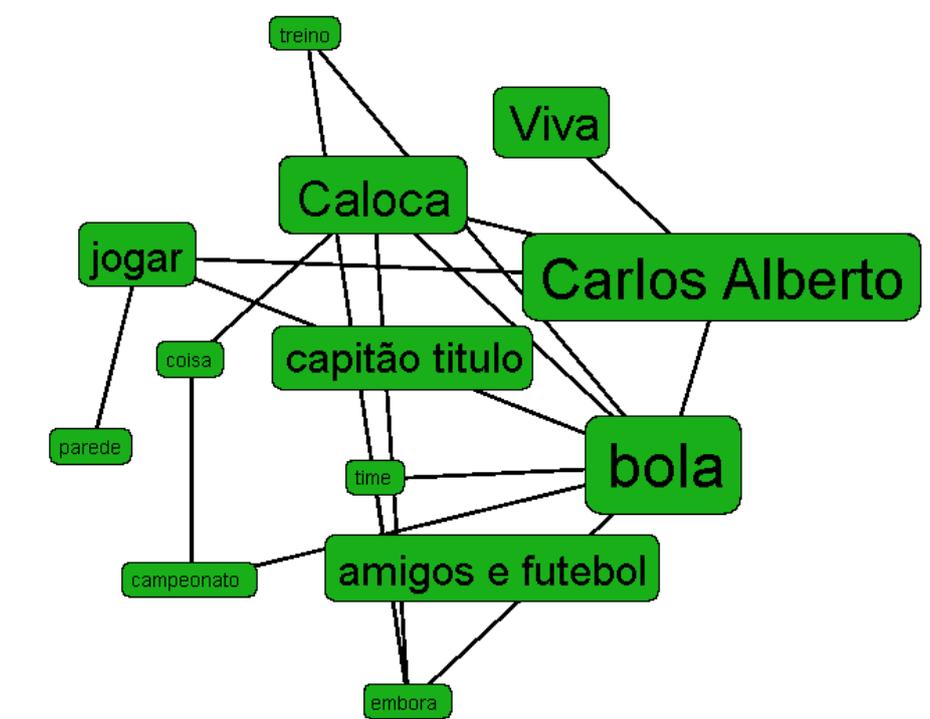


Figura 17 Grafo referente ao texto “O dono da bola” e alterado pelo estudante A

Os estudantes de primeiro ciclo tiveram um pouco mais de dificuldades na utilização da ferramenta que os estudantes de terceiro ciclo. Não apenas o componente tecnológico influenciou mais o resultado da atividade, como também evidenciou em que medida a capacidade dos alunos de inserir/remover termos do grafo foi um fator relevante. A figura 18 apresenta um apanhado dos grafos produzidos por diversos alunos de primeiro ciclo.

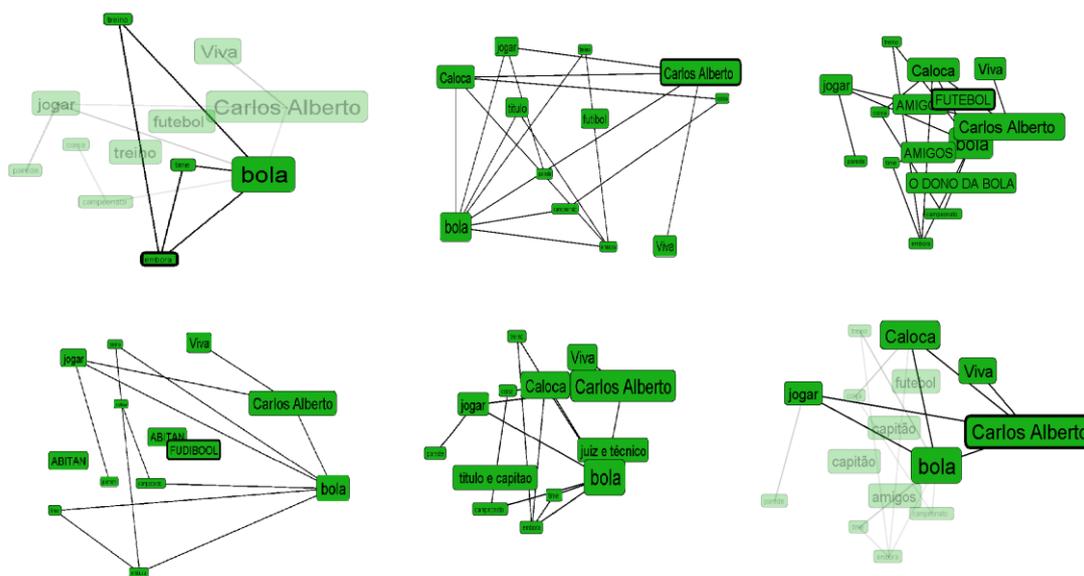


Figura 18 Grafos construídos pelos estudantes de primeiro ciclo para o texto "O dono da bola"

Ao analisar os dados coletados da interação dos alunos do Ciclo 3, percebe-se que o número de adição de nodos é quase idêntico ao de remoção. Além disso, a ação inicial destes alunos é a de remoção de nodo, ao contrário do que ocorre com o Ciclo 1. As ações de edição de nodo também são bastante frequentes, demonstrando que os novos termos adicionados não permanecem isolados dos demais. No geral, há um número proporcionalmente maior de ações realizadas no grafo por estes estudantes que aquele contabilizado para estudantes do outro ciclo.

As informações coletadas demonstram que os estudantes do ciclo 3 moldam o grafo à sua percepção do texto, tendo mais confiança e segurança no que estão fazendo. Eles não apenas adicionam conteúdo, mas também os relacionam com aqueles já existentes e alteram as relações criadas pelo Sobek. O grafo inicial é tomado por base pelos alunos, que procuram remover e adicionar termos de forma a manter a estrutura semelhante àquela de início. Há muitas conexões refeitas, o que demonstra que estes estudantes compreendem melhor o significado das arestas no grafo e têm maior capacidade de relacionar informações.

Os resultados permitem concluir que eles rapidamente compreenderam a atividade e a ferramenta, além de possuírem um maior nível de confiança nas suas ações e melhor reflexão sobre o texto e sobre o grafo que os estudantes do ciclo 1. Ao

remover e adicionar nodos e arestas, os estudantes trazem representações mais fiéis de suas estruturas cognitivas para o grafo como um todo, não apenas avaliando termos isolados. O processo de construção e desconstrução do conhecimento presente no grafo reforça os conceitos de aprendizagem significativa, gerando modificações derivativas e correlativas no corpo de conhecimento do aluno. Ao observar o grafo e os relacionamentos entre termos, o estudante tem a chance de refletir se aquelas conexões estão de acordo com o que ele conhece e se há informações que possam ser relacionadas ao conhecimento que já possui. Se ele decidir por remover um relacionamento, por exemplo, ele está utilizando seu conhecimento prévio que indica que não há uma ligação entre dois termos, reforçando seu conhecimento sobre tais conceitos.

Os textos apresentados aos estudantes do ciclo 3 eram mais complexos que aqueles do ciclo 1 e os grafos apresentados pelo Sobek possuíam um menor número de termos/relações, dando aos estudantes maior margem para realização de alterações e reconstruções. O primeiro texto, chamado “*Defenestração*”, de Luis Fernando Veríssimo, produziu um grafo com termos compostos, termos isolados dos demais, além de termos que não trazem (de forma individual) informação ao usuário, tais como advérbios. O grafo pode ser visualizado na Figura 19.

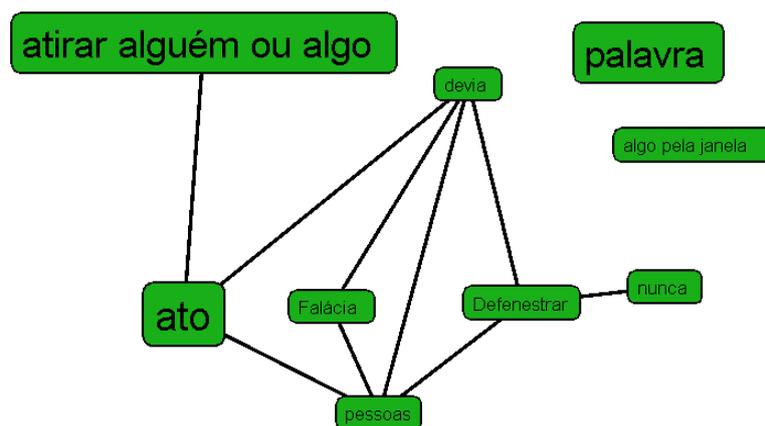


Figura 19 Grafo construído pelo Sobek a partir do texto de Luis Fernando Veríssimo

A partir deste grafo inicial, os estudantes criaram representações diferentes do texto. Uma coletânea de grafos produzidos pelos estudantes pode ser visto na figura 20.

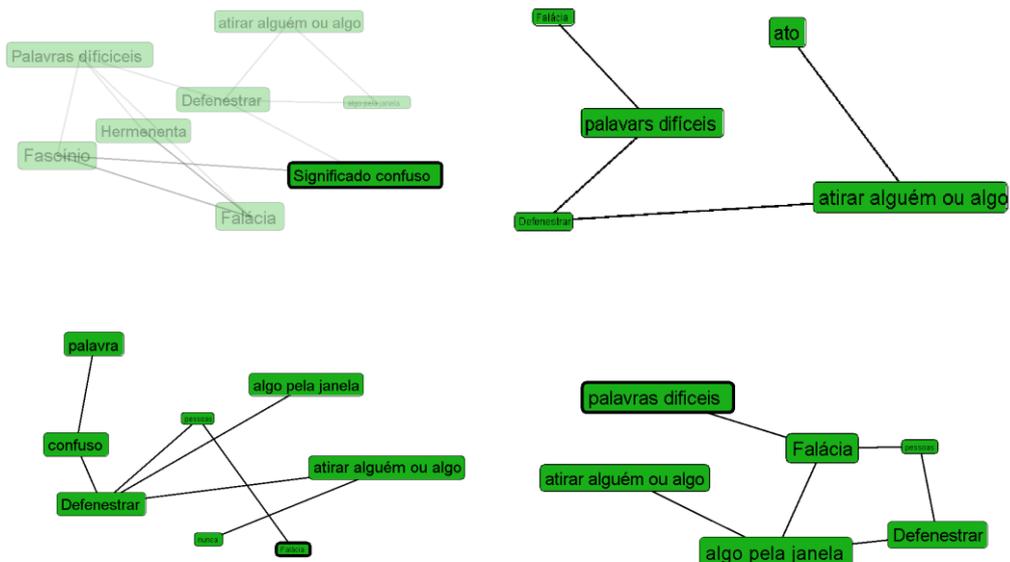


Figura 20 Grafos construídos pelos estudantes de terceiro ciclo para o texto "Defenestrar"

Foi realizada uma análise dos grafos produzidos por dois estudantes: o estudante B, que acertou o maior número de questões nos textos, e o estudante C, que obteve o menor número de acertos. Os grafos criados por cada um dos estudantes podem ser visualizados nas figuras 21 e 22.

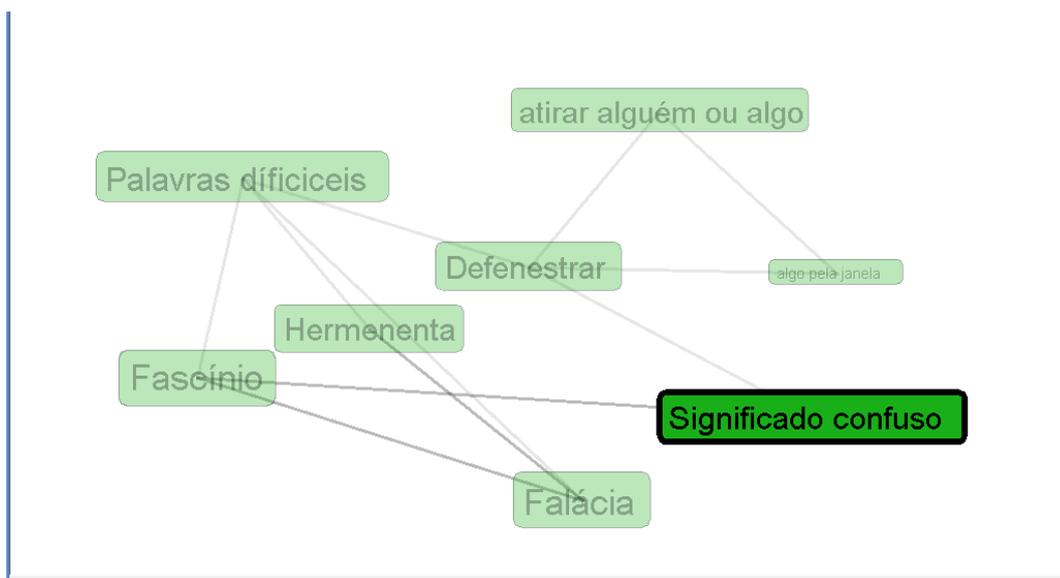


Figura 21 Grafo produzido pelo estudante B, estudante com melhor desempenho no terceiro ciclo

É possível observar que o grafo produzido pelo *estudante B* é bastante diferente daquele gerado pelo Sobek. Muitos termos foram adicionados ao grafo e outros tantos removidos. Esse conjunto de ações permite inferir que o estudante fez um grande investimento na tentativa de refletir e compreender o texto, uma vez que identificou como tornar o grafo mais representativo através do conjunto de alterações realizado. Considerando os relacionamentos criados entre nodos, a análise do grafo indica que o estudante identificou corretamente tanto o significado dos termos (nos relacionamentos entre o termo “*defenestrar*”, “*atirar alguém ou algo*” e “*jogar pela janela*”) quanto os relacionamentos presentes no texto (nos relacionamentos entre “*Palavras difíceis*”, “*Defenestrar*” e “*Hermeneuta*”). A compreensão do texto pelo aluno pode ser observada no grafo por ele construído, indicando os termos considerados principais para a representação. Esse entendimento e ação sobre o conhecimento pode explicar o fato de este estudante ter apresentado o melhor desempenho dentre os alunos do ciclo 3.

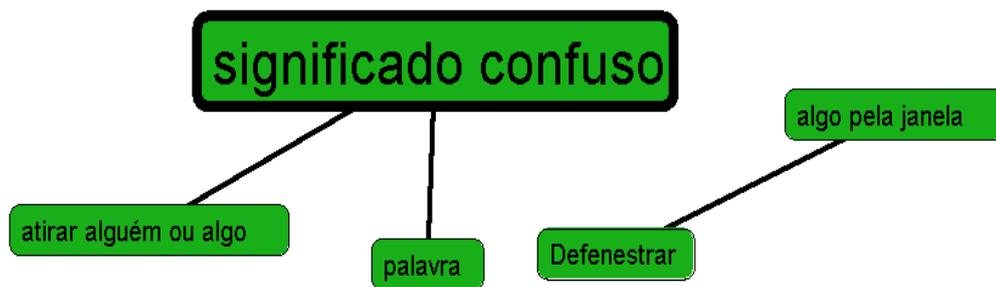


Figura 22 Grafo produzido pelo *estudante C*, estudante com menor número de acertos no texto de Luis Fernando Veríssimo

Por outro lado, o *aluno C* apresentou o pior desempenho entre todos os estudantes em termos de número de questões corretamente respondidas. O grafo produzido por ele apresenta um único termo adicionado, “*significado confuso*”. Muitos termos foram retirados do grafo, mas não se observam termos novos inseridos. Os termos presentes apontam para a compreensão e reflexão sobre aspectos pontuais do texto, mais precisamente sobre o aspecto relacionado ao termo “*defenestrar*”. Os nodos contêm informações sobre apenas alguns parágrafos do texto e os relacionamentos presentes no grafo não representam claramente a ligação de ideias presentes no texto. Os nodos “*significado confuso*” e “*atirar alguém ou algo*” não estão relacionados por significado ou por relação dentro do texto, o que torna o par um candidato pouco comum para ter um relacionamento.

As alterações no grafo denotam o comportamento de um estudante que removeu aquilo que não compreendia (como o termo “*falácia*”, que é relevante para o contexto do texto e não aparece em seu grafo final) e aquilo que julgou não estar presente no texto. Estas ações podem estar relacionadas ao fato do estudante não ter sido capaz de dar significado ao conteúdo do texto. Por mais que possa haver uma reflexão sobre o texto, não há aparente compreensão do assunto principal do texto e não foram adicionados termos capazes de representá-lo de forma completa ou mesmo parcial. Esse resultado pode não estar diretamente relacionado à capacidade de leitura do aluno, mas sim à ausência de subsunçores que permitissem a ele dar significado ao que foi lido. Ademais, apenas o relacionamento entre os termos “*defenestrar*” e “*algo pela janela*” parecem estar de acordo com o conteúdo do texto e com o significado dos termos. Aparentemente, o estudante apenas focou na identificação do assunto considerado central no texto, sem aprofundar sua compreensão sobre o que o texto tratava ou porque os termos apresentados pela ferramenta estavam em destaque.

Observando as diferenças entre os dois grafos e o resultado da avaliação dos alunos, é possível inferir que o estudante que se empenhou mais em organizar e produzir um grafo mais completo e coeso (*estudante B*) compreendeu melhor o texto e encontrou as informações que nele estavam implícitas. Já o *estudante C* não trabalhou sobre o grafo ou não conseguiu manipular a ferramenta de forma adequada. Essa falta de ação sobre o conhecimento pode ter prejudicado a compreensão do aluno sobre os fatos apresentados no texto, resultando em um baixo desempenho na avaliação.

Conforme descrito no Capítulo 3, um dos principais desafios do contexto escolar é encontrar ferramentas e estratégias de ensino que incentivem o estudante a agir sobre o conhecimento e realizar as ações necessárias para aprender de forma significativa. No caso do *estudante C*, especificamente, a ferramenta de mineração Sobek pode não ter instigado o aluno suficientemente para levá-lo a refletir sobre o texto e realizar adequadamente as ações propostas na atividade para auxiliá-lo na compreensão do texto. Por se tratarem de textos complexos (como pode ser visto na análise das respostas dos estudantes na sessão 6.3), a falta de reflexão e ação sobre o conhecimento presente no texto resultou em uma interpretação indevida, trazendo como consequência um baixo número de acertos no atividade de avaliação.

### 6.3 Questionário com estudantes participantes da pesquisa

A última etapa da intervenção realizada com alunos do terceiro ciclo consistiu em um questionário onde eles puderam expressar suas opiniões acerca do Sobek e da atividade. Semelhante ao questionário aplicado aos estudantes no estudo preliminar, esse questionário continha 6 perguntas objetivas e 4 perguntas dissertativas.

As perguntas com opções objetivas foram:

1. Qual o seu nível de satisfação em relação às atividades realizadas?
2. Qual o seu nível de satisfação em relação à ferramenta Sobek?
3. Você considera que o grafo representa o texto?
4. Qual o grau de dificuldade para utilizar a ferramenta Sobek?
5. Nas atividades de leitura, você considera que o Sobek ajudou a entender o texto?
6. Você utilizaria o Sobek fora da sala de aula?

Com o intuito de padronizar as opções de resposta, as perguntas de 1 à 4 possuíam 5 opções para o aluno escolher, sendo a primeira opção uma afirmação de teor positivo quanto a questão (exemplo: “*considerarei muito bom*”) e a última opção uma afirmação de teor negativo para a questão (exemplo: “*considerarei muito ruim*”).

As duas primeiras questões apresentadas representam a satisfação geral dos estudantes quanto à atividade e à ferramenta. Para a primeira pergunta, 75% dos alunos (9 de 12) apontaram que “*gostaram ou gostaram muito*” da atividade e apenas um aluno respondeu que “*não gostou*” (Figura 23). O resultado corrobora com o alcançado no estudo preliminar, demonstrando que a atividade estava bem elaborada e foi agradável aos estudantes. Conforme apontado pela professora apoiadora do estudo preliminar, o formato e conteúdo multimídia da atividade incentivaram os estudantes a realizarem ela com mais disposição.

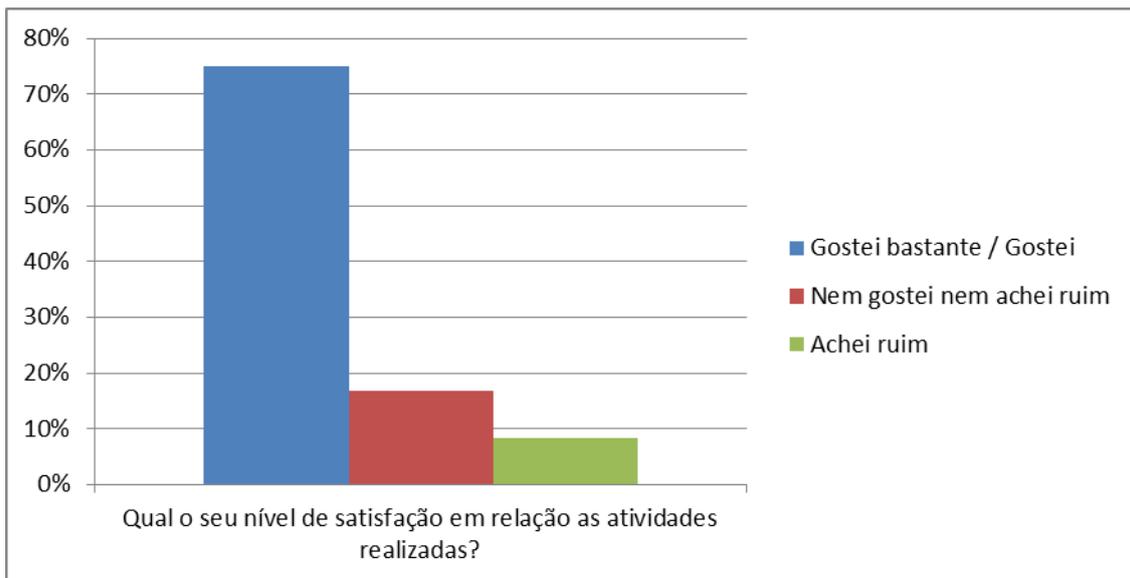


Figura 23 Respostas dos estudantes a questão 1 do questionário de avaliação da atividade

O resultado obtido na segunda pergunta apresentou um nível de satisfação de 58,3%, um valor abaixo daquele obtido previamente. Mais de metade dos estudantes aprovaram a ferramenta Sobek e apenas dois alunos responderam que não gostaram do minerador. Um ponto a destacar é que ninguém atribuiu o valor mínimo para o Sobek. Certamente, há pontos a serem aprimorados na ferramenta, mas de forma geral a maioria aprovou o software e poucos reprovaram o seu uso.

Apesar de o Sobek ter obtido um resultado menos positivo na avaliação dos estudantes, 66,7% deles considerou que o Sobek representou bem ou muito bem o texto minerado. Considerando os resultados obtidos nas perguntas 2 e 3 (Figura 24), podemos afirmar que o algoritmo de mineração do Sobek é apropriado e que sua interface e usabilidade implementam a função para o qual foi concebido. Pode-se também considerar que o algoritmo de mineração da ferramenta necessita de menos alterações que sua interface gráfica para que o usuário obtenha uma experiência agradável e resultados confiáveis.

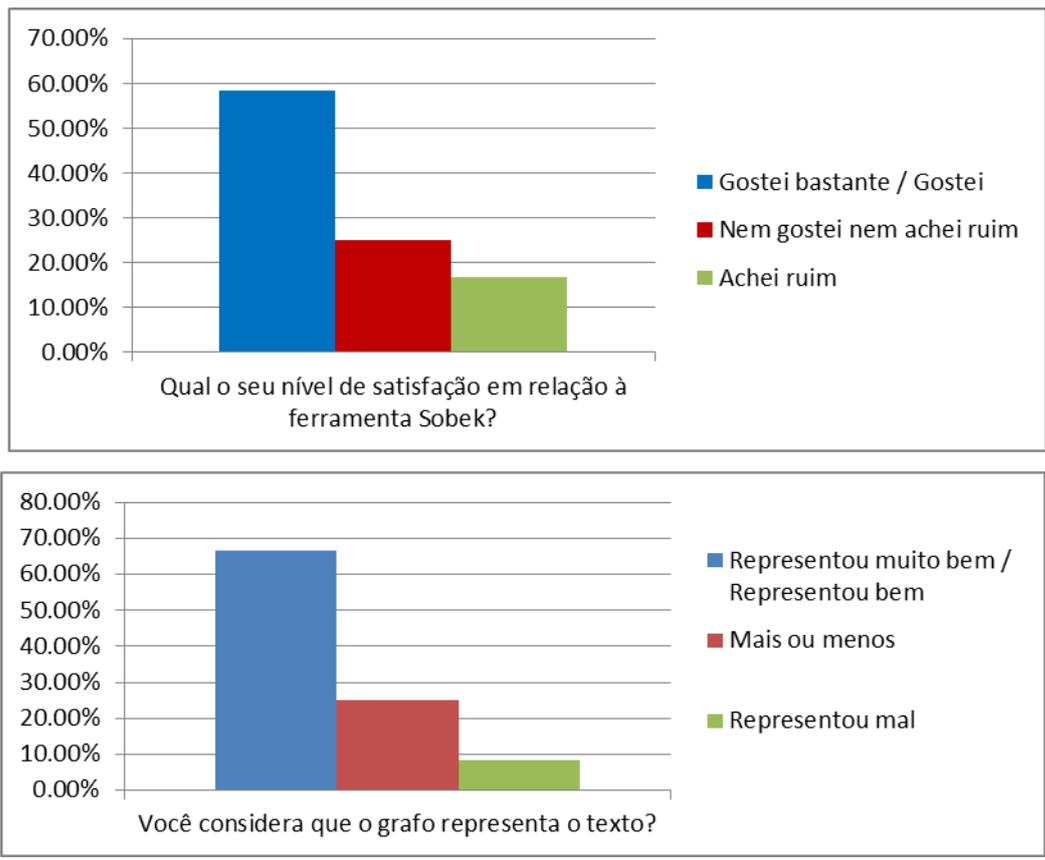
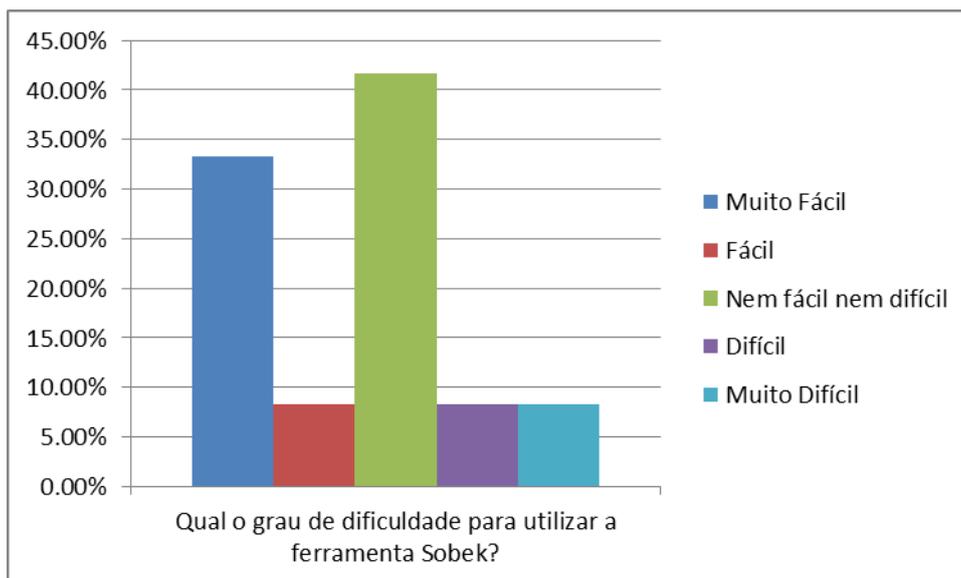


Figura 24 Resultados obtidos das respostas das questões 2 e 3

Um quesito que requer atenção e um estudo mais aprofundado é o nível de dificuldade no uso da ferramenta. No estudo preliminar, essa dificuldade foi diagnosticada nas intervenções iniciais e a ferramenta se tornou gradativamente mais conhecida e assimilada pelos estudantes. No caso do estudo apresentado, os estudantes tiveram apenas duas intervenções com a ferramenta, uma demonstrativa e uma segunda onde ela já foi utilizada na prática. Dessa forma, não é surpresa que a pergunta 4 apresente grande parte dos alunos considerando que a ferramenta “*não é fácil nem difícil de usar*” e até alunos que a consideraram muito difícil (Figura 25). Apesar da simplicidade da interação com a ferramenta, a atividade também exigiu que os estudantes refletissem sobre termos e relacionamentos, tarefa que pode ter se tornado mais complicada tendo em mãos uma tecnologia que até então não lhes era familiar.



**Figura 25** Gráfico representando a dificuldade dos estudantes no uso da ferramenta Sobek

É importante analisar o grau de dificuldade de manuseio da ferramenta Sobek com bastante atenção. A ferramenta foi desenvolvida para ser utilizada sem a necessidade de treinamento prévio e por professores e estudantes com pouca experiência na área tecnológica. Tendo em vista o grande número de estudantes que não considerou a ferramenta como de fácil manuseio, torna-se necessário identificar qual ação sobre o minerador está criando esse efeito que fez com que os estudantes não considerassem a ferramenta de fácil uso. Mesmo que o Sobek tenha sido desenvolvido com enfoque pedagógico, é possível que ele ainda precise sofrer alterações para que possa ser utilizado amplamente em sala de aula.

Na questão 5, é confirmado que a boa representação do texto pelo Sobek auxilia os estudantes a compreenderem o texto; 75% dos estudantes responderam que sim, o Sobek ajudou a entender o texto. Esse resultado também corrobora aqueles obtidos pela avaliação qualitativa, que demonstrou que de fato os estudantes responderam mais questões corretamente quando apoiados pela ferramenta. Nenhum participante respondeu que o Sobek não ajudou a compreender o texto.

Por fim, a maioria dos estudantes respondeu a questão 6 afirmando que não utilizaria o Sobek fora da sala de aula. Isso não está relacionado com uma desaprovação da ferramenta (tendo em vista que a grande maioria respondeu que “gostou” ou “gostou muito” dela), mas sim com a forma com que ela foi disponibilizada. Muitos estudantes

tiveram dificuldade em utilizar o minerador (principalmente nos estágios iniciais da intervenção). O fato de não ter sido disponibilizado uma versão para dispositivos móveis pode ter sido determinante nas respostas dos alunos.

Após responderem às seis perguntas objetivas, os estudantes tiveram a oportunidade de responder mais 4 perguntas dissertativas. As perguntas foram:

7. Quais os pontos positivos do Sobek?
8. Quais os pontos negativos do Sobek?
9. Como você acha que o Sobek pode melhorar?
10. O que você achou das atividades?

De forma geral, o ponto positivo mais destacado do minerador de textos Sobek foi o seu auxílio para compreensão textual. A maioria dos estudantes escreveu que o minerador os ajudou a entender os textos. Este é um dos objetivos do desenvolvimento deste trabalho e a visão dos estudantes valida a noção de que esta tese atingiu os objetivos propostos. Os demais pontos positivos elencados pelos estudantes são voltados ao apoio dado pela ferramenta em algum aspecto específico da compreensão textual. Diferente da avaliação realizada durante o estudo preliminar, características de interação e design da ferramenta não foram mencionadas desta vez. A dificuldade encontrada por diversos alunos pode explicar porque estes pontos não aparecem na avaliação positiva do Sobek.

Por outro lado, na avaliação negativa da ferramenta e na questão acerca dos pontos a melhorar do Sobek, diversos alunos mencionaram que acharam difícil interagir com ele, principalmente na edição do grafo. Alguns alunos consideraram que o grafo inicial apresentado (com os termos ainda embaralhados) era um pouco confuso para entender e que seria importante que a apresentação fosse mais simples. Outros não elencaram pontos negativos nem aspectos a melhorar, tanto por não terem pensando em “*nenhuma ideia para escrever*” quanto por acharem que não havia pontos negativos ou a melhorar. Não foram mencionadas questões relacionadas com letramento, apenas com design e interface.

A última questão presente foi relacionada à intervenção realizada. A maioria das respostas qualificou a atividade como “*legal*” e “*complexa*”. De acordo com o retorno dos alunos, os textos escolhidos para a atividade com o Sobek eram bastante complexos

e difíceis de compreender. Apesar disso, a atividade é classificada como interessante pela maioria e é comentado nas respostas que o Sobek ajudou a responder as perguntas do texto.

As perguntas do questionário apresentado aos estudantes avaliam tanto a facilidade de uso percebida do sistema, quanto a percepção dos estudantes com relação à utilidade das atividades propostas. A facilidade de uso percebida de um artefato tecnológico é um fator determinante para sua utilização efetiva (Venkatesh et al., 2003), o que de certo modo aponta para a necessidade de reavaliar determinados aspectos relativos à construção da interface do Sobek. Mesmo assim, o fato de os alunos terem tido uma percepção positiva com relação ao interesse da atividade para a compreensão textual demonstra o potencial da ferramenta Sobek no apoio à aprendizagem relacionada à leitura.

## 7. Conclusão

O estudo realizado nesta tese buscou investigar como o minerador de textos Sobek poderia apoiar os estudantes no processo de leitura e compreensão textual. A ferramenta é um minerador de textos que extrai os termos mais relevantes encontrados em um texto e os representa na forma de um grafo. Ele foi inicialmente desenvolvido para apoiar professores na tarefa de sumarização de textos em ensino à distância (Macedo, 2009), mas suas possibilidades de uso fizeram o projeto investigar a sua aplicabilidade em também apoiar estudantes nos processos relacionados ao letramento. Sua abordagem está alinhada com a pesquisa de aprendizagem significativa, formulada por Ausubel (1968) e com inspiração nas pesquisas de mapas conceituais de Novak (2002). O grafo provido pelo minerador auxilia tanto na identificação de ideias importantes no texto quanto no estabelecimento de relacionamentos entre elas. Ao interagir com o grafo, o usuário põe em prática um elemento fundamental para a aprendizagem significativa que é sua postura ativa frente à atividade de aprendizagem e seu interesse em aprender. A representação do texto na forma gráfica auxilia na aprendizagem pela possibilidade de visualização dos relacionamentos entre conceitos e ideias, favorecendo principalmente estudantes com baixa capacidade de leitura e compreensão textual (Manoli e Papadopoulou, 2012). Acredita-se que isto contribui para a significância de seu aprendizado.

Foi postulado que o minerador de textos Sobek poderia auxiliar estudantes no processo de compreensão textual. Considerando os resultados apresentados pela Avaliação Nacional da Alfabetização no ano de 2014, em que mais da metade dos estudantes do terceiro ano é incapaz de ler textos longos ou extrair informações que não estejam explicitamente descritas neles, foi considerado fundamental abordar tal questão e desenvolver ferramentas que pudessem apoiar os estudantes no processo de leitura e compreensão de textos. Ao extrair termos relevantes e fornecer uma representação não linear de um texto, a ferramenta Sobek auxilia os estudantes a refletir sobre a leitura, buscar informações que conhecem e relacionar estes fatos com o que foi lido, encontrar informações no texto e compreender o relacionamento entre termos, auxiliando justamente nos aspectos mais destacados do resultado da ANA.

Neste trabalho, dois experimentos foram realizados para validar a proposta do pesquisador. O primeiro experimento, chamado no texto de pesquisa preliminar, foi realizado em uma escola pública de Porto Alegre com alunos que apresentavam baixo desempenho em questões relacionadas ao letramento. Os estudantes participantes das atividades possuíam dificuldades para compreensão e produção textual e faziam parte de um grupo de apoio escolar disponibilizado pela escola. Estas atividades serviram para que fosse possível avaliar o uso do Sobek por estudantes que fazem parte do público alvo da ANA.

O estudo da fase preliminar também serviu para coletar a percepção dos estudantes quanto ao Sobek e às atividades. Durante a avaliação com alunos e com a professora responsável, o minerador de texto foi elogiado e reconhecido como uma ferramenta de apoio, principalmente do ponto de vista de leitura e compreensão de texto. Os estudantes participantes da pesquisa gostaram de utilizar a ferramenta e argumentaram que ela auxiliou na compreensão daquilo que foi lido. Também destacaram que gostariam de utilizar ela em diferentes ocasiões. A professora responsável por acompanhar a pesquisa na escola também enfatizou que a ferramenta poderia ser utilizada em sala de aula e que teria um melhor resultado na medida em que os textos trabalhados fossem mais complexos.

Durante as intervenções desse primeiro experimento, foram identificadas questões relacionadas à ferramenta e à atividade que poderiam ser aprimoradas. Com base nos dados coletados, o minerador de textos sofreu alterações no código para se tornar mais rápido e coletar de forma automática as ações dos usuários, para serem posteriormente analisadas. Estas alterações foram colocadas em prática no experimento final realizado para esta tese. As observações quanto ao uso da ferramenta e teor das atividades também foram utilizadas para aprimorar a seleção de textos e a disponibilização da ferramenta para os estudantes.

O experimento final (conclusivo) deste trabalho foi realizado em uma escola de Caxias do Sul, com estudantes participantes de um projeto chamado “*Ação Social*”. O projeto é voltado aos estudantes considerados “*em situação de risco e vulnerabilidade social*”. Estes estudantes apresentam defasagem em aspectos pedagógicos, principalmente em questões relacionadas ao letramento (a exemplo do conjunto de estudantes da pesquisa preliminar). No total, 43 estudantes de dois ciclos do projeto

participaram da pesquisa. Os estudantes realizaram atividades de leitura e interpretação textual com e sem o apoio da ferramenta Sobek.

Os resultados obtidos no estudo demonstram que o Sobek foi capaz de auxiliar os estudantes nos aspectos de compreensão textual e interpretação de textos. Nas intervenções em que ele foi utilizado, os estudantes conseguiram, em média, responder um maior número de questões corretamente. Estes resultados foram analisados de forma estatística, demonstrando que há uma correlação positiva entre o uso do minerador e o aumento no número de questões respondidas corretamente, principalmente no que concerne os estudantes do 3o ciclo. Foi também demonstrado que os dados apresentados são estatisticamente significativos.

A análise dos grafos produzidos pelos estudantes permitiu identificar que os alunos de primeiro ciclo tiveram como ação principal adicionar termos ao grafo. Um baixo número de alunos removeu nodos ou editou suas relações. A premissa de que o minerador de textos Sobek é capaz de encontrar termos relevantes e apresentá-los na forma de grafo pode ter inibido a ação de remoção de termos, pois ela exige que o usuário confronte uma ideia apresentada a ele pela ferramenta. Por outro lado, estudantes de terceiro ciclo removeram e adicionaram termos em proporção semelhante. Estes estudantes são mais confiantes quanto a sua capacidade de ler e compreender textos e possuem mais experiência em tal atividade que os de ciclo 1. Dessa forma, se sentiram mais confortáveis em moldar o grafo ao que compreendiam ser o enfoque principal do texto.

Avaliando os termos adicionados ao grafo, foi constatado que termos que não estavam presentes no texto eram adicionados pelos estudantes e conectados a termos já existentes. Isso demonstra que os estudantes refletiram sobre o texto e trouxeram seus próprios conhecimentos para enriquecer a representação textual proposta pelo Sobek. Considerando a teoria de aprendizagem significativa de Ausubel (1978), a intervenção permitiu que os estudantes relacionassem o conhecimento que possuíam com as informações novas apresentadas a eles. Tal ação somada à vontade de aprender e interagir com a ferramenta são pontos fundamentais para afirmar que a intervenção, através da ferramenta, fornece os subsídios para aprendizagem significativa do aluno quando considerado o conteúdo do texto. Também permite afirmar que o estudante

refletiu sobre o texto, pois foi capaz de relacionar as informações com os conhecimentos prévios que possuía e que não estavam presentes no texto.

Também era objetivo específico desta pesquisa uma avaliação da ferramenta e das atividades por meio de questionário disponibilizado aos estudantes. Este questionário continha 6 questões objetivas e 4 questões dissertativas. As respostas indicaram que o Sobek é uma ferramenta que os estudantes gostaram de utilizar e que as atividades estavam de acordo com o esperado. Os estudantes de ambos os estudos demonstraram aprovar a ferramenta e as atividades, de maneira geral.

Uma das questões respondidas durante a pesquisa versava sobre a dificuldade de utilização do Sobek. Foi possível identificar que a ferramenta não se apresentou tão simples quanto o esperado. Mesmo os estudantes tendo conseguido utilizá-la e desenvolver os grafos de acordo com o requisitado, muitos deles comentaram sobre a complexidade da proposta e sobre dificuldades de uso.

Um fator que pode explicar tais opiniões é a terminologia utilizada pela ferramenta. Ficou evidente durante todas as atividades que os termos utilizados para descrever a interação do usuário com o Sobek são demasiados restritos e não são de conhecimento dos estudantes. Mesmo o conceito de “*grafo*” (termo central ao Sobek), apesar de semanticamente correto, não é conhecido pela maioria dos estudantes. Um exemplo disso ocorreu durante a segunda intervenção executada na escola em Caxias, quando os estudantes corrigiram o colega que falou “*grafo*”, dizendo para ele que o correto seria “*gráfico*”. Os termos “*nodo*” e “*extração de termos*” também foram constantemente questionados. A maioria dos estudantes se referia aos nodos como “*bolinhas*” ou “*caixinhas*”. Um dos participantes chegou a escrever em sua resposta à questão que indagava sobre os pontos positivos do Sobek que “*É muito legal ficar brincando com as bolinhas*”. Pode-se, portanto, constatar que a terminologia empregada pode inibir os estudantes e dificultar o uso da ferramenta e exploração de todos os seus recursos.

Um segundo aspecto, evidenciado durante as intervenções, é a necessidade de desenvolver a ferramenta Sobek para plataformas móveis. O uso do computador pelos alunos (principalmente aqueles mais jovens e da escola pública de Porto Alegre) é bastante limitado, sendo que muitos deles não utilizam o computador fora da escola. Por

outro lado, a destreza na manipulação de dispositivos móveis demonstra que o Sobek pode vir a ser mais utilizado se fosse disponibilizado para tais dispositivos.

Em vista das observações realizadas durante as intervenções e dos comentários dos estudantes, foi iniciado o desenvolvimento de uma versão para dispositivos móveis da ferramenta. Essa versão possui uma interface de manipulação de grafos bastante semelhante àquela da versão para computadores de mesa. Dessa forma, é possível adequar as pesquisas realizadas nesse trabalho para embasar o desenvolvimento dessa nova versão da ferramenta.

Mesmo a ferramenta sendo utilizada pelos estudantes apenas na sua versão para computador, alguns alunos projetaram utilizar a ferramenta em contextos específicos. Um deles comentou que “... o Sobek é bom para resumo, mas não serve para estudar para provas”. Este mesmo aluno demonstrou interesse em utilizar o Sobek fora da sala de aula, mostrando que viu potencial em poder resumir textos ou mesmo compreender eles de forma mais rápida. Outro aluno perguntou se “era possível colocar imagens nos termos”. A possibilidade de extração de termos com imagens vinculadas aos nodos já foi contemplada pelo pesquisador e foi implementada para o trabalho de Langa (2014), onde foi estudada a contribuição da ferramenta utilizando no processo de alfabetização.

Considerando as respostas dadas pelos alunos para as perguntas dissertativas, é possível perceber que os estudantes realizaram uma análise crítica sobre a ferramenta e a utilização dela. Muitas respostas e comentários individuais contemplam a capacidade do minerador e as diferentes formas de uso. Respostas como “*Descobrimos uma ferramenta para nos ajudar a compreender os textos*”, “*Acho que melhor impossível porque o intuito é de fazer as pessoas pensarem de uma forma diferente para o texto, não dar a resposta sem ler o texto*” e “*(O Sobek nos ajuda a) Entender as palavras difíceis no texto, entender a relação do texto e das palavras*” demonstram que o objetivo da ferramenta e a sua utilização são claramente identificados pelos estudantes. Outras respostas como “*Os textos são um pouco complexos, porém, um estímulo intelectual*” e “*No começo achei os textos confusos...*” indicam que os textos da intervenção realizada com o Sobek não são triviais e que o minerador auxiliou os estudantes a responderem corretamente o questionário, facilitando a compreensão do texto.

Considerando os resultados obtidos, é possível afirmar que o objetivo da pesquisa foi atingido de forma satisfatória. Também é importante ressaltar a forte aprovação que a ferramenta obteve, tanto de professores quanto de estudantes. Estes resultados demonstram que o minerador de textos Sobek pode ser utilizado para apoio aos processos de letramento, sendo especificamente formalizado a sua contribuição para aspectos relativos à compreensão de textos.

Como trabalhos futuros, além da já iniciada versão para dispositivos móveis, a questão da terminologia utilizada no Sobek deve ser analisada. Outro aspecto que deve continuar a ser explorado é o uso da ferramenta para a produção textual. Durante este trabalho, foi utilizado o grafo produzido pelos alunos ou extraído de textos para que o escritor encontrasse subsídios para escrever e também com o objetivo do aluno organizar as ideias que devem estar presentes no texto antes de escrever ele. O enfoque no auxílio a professores também está sendo desenvolvidos. É proposto que se possa utilizar o minerador de textos Sobek para buscar informações sobre a coesão e coerência de textos. Utilizando o próprio texto dos estudantes, o Sobek poderia encontrar relacionamentos entre as diferentes partes do texto e comparar os termos mais recorrentes com aqueles propostos na produção textual. Este processo contaria com novas funcionalidades a serem implementadas na ferramenta, como a capacidade de comparação entre textos onde o resultado seria o conjunto de termos iguais (ou diferentes) presentes em ambos os textos.

Outra importante funcionalidade a ser desenvolvida e testada é o uso de analisadores de classe gramatical na mineração de textos. A pesquisa nesse tema pode encontrar resultados que tornariam o grafo apresentado pelo Sobek mais útil para os estudantes e demais usuários da ferramenta. Seria possível destacar termos de determinada classe gramatical, remover eles ou apresentar apenas os termos que correspondam à classe requerida. O resultado da mineração para estes casos deve ser reavaliado, pois seria necessário que o algoritmo do Sobek levasse em consideração as restrições necessárias ao fazer a mineração. Porém, já está sendo pesquisado se classes gramaticais específicas (tais como verbos) devem aparecer no grafo ou se deveriam ser completamente removidas. Grafos mais restritos podem ser mais representativos do texto ou mesmo podem servir melhor de base para produção textual.

# Referências

- Acosta, O., Behar, P., and Berni Reategui, E. (2014). Content recommendation in an inquiry-based learning environment. In *Frontiers in Education Conference (FIE), 2014 IEEE*, pages 1–6.
- Adams, M. (1994). *Beginning to Read: Thinking and Learning about Print*. Bradford books. BRADFORD BOOK.
- Adams, M. J. (2011). The relation between alphabetic basics, word recognition, and reading. In S. Jay Samuels, A. E. F., editor, *What Research Has to Say About Reading Instruction*, pages 4–25. International Reading Assoc.
- Adams, M. J., Treiman, R., and Pressley, M. (1998). Reading writing, and literacy. In  
D. Kuhn, R. S. E., editor, *Cognitive Language, and Perceptual Development (Vol. 2)*. Wiley.
- Anderson, R. C. and Pearson, P. D. (1988). A schema-theoretic view of basic processes in reading comprehension. In Carrell, P. L., Devine, J., and Eskey, D. E., editors, *Inter- active Approaches to Second Language Reading*, pages 37–55. Cambridge University Press. Cambridge Books Online.
- Anderson, V. (1992). A teacher development project in transactional strategy instruction for teachers of severely reading-disabled adolescents. *Teaching and Teacher Educa- tion*, 8(4):391 – 403.
- Ausubel, D. P. (1963). *The psychology of meaningful verbal learning*. Grune & Stratton.
- Ausubel, D., Novak, J., and Hanesian, H. (1978). *Educational Psychology: A Cognitive View*. Holt, Rinehart and Winston.
- Azevedo, B. F., Reategui, E., and Behar, P. A. (2014). Analysis of the relevance of posts in asynchronous discussions. *Interdisciplinary Journal of Knowledge and Learning Objects*, 10:107–121.
- Bachelard, G. and Godinho, F. (1971). *Epistemologia*. Edição 70.
- Baron, J. (1977). Mechanisms for pronouncing printed words: Use and acquisition.

- In (Eds), D. L. . S. J. S., editor, *Basic processes in reading: perception and comprehension*, pages 175–216. Hillsdale, NJ: Erlbaum.
- Baron, J. and McKillop, B. J. (1975). Individual differences in speed of phonemic analysis, visual analysis, and reading. *Acta Psychologica*, 39(2):91 – 96.
- Beck, I. L., Perfetti, C. A., and McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 74:500–512.
- Beers, K. (2003). *When kids can't read. What teachers can do*. Portsmouth, NH: Heinemann.
- Beissner, K. L., Jonassen, D. H., and Grabowski, B. L. (1994). Using and selecting graphic techniques to acquire structural knowledge. *Performance Improvement Quarterly*, 7(4):20–38.
- Belardinelli, M. O., Palmiero, M., Sestieri, C., Nardo, D., Matteo, R. D., Londei, A., D'Ausilio, A., Ferretti, A., Gratta, C. D., and Romani, G. (2009). An fmri investigation on image generation in different sensory modalities: The influence of vividness. *Acta Psychologica*, 132(2):190 – 200.
- Block, C. C. and Pressley, M. (2002). *Comprehension Instruction: Research-based Best Practices*. Solving problems in the teaching of literacy. Guilford Press.
- Bolasco, S., Canzonetti, A., Capo, F., della Ratta-Rinaldi, F., and Singh, B. (2005). Understanding text mining: A pragmatic approach. In Sirmakessis, S., editor, *Knowledge Mining*, volume 185 of *Studies in Fuzziness and Soft Computing*, pages 31–50. Springer Berlin Heidelberg.
- Breznitz, Z. (1997). Effects of accelerated reading rate on memory for text among dyslexic readers. *Journal of Educational Psychology*.
- Brown, A. L. and Day, J. D. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior*, pages 1–14.
- Calvo, R. A., O'Rourke, S. T., Jones, J., Yacef, K., and Reimann, P. (2011). Collaborative writing support tools on the cloud. *IEEE Transactions on Learning Technologies*, 4(1):88–97.
- Carletta, J. (1996). Assessing agreement on classification tasks: The kappa

- statistic. *Comput. Linguist.*, 22(2):249–254.
- Chang, K. E., Sung, Y. T., and Chen, S. F. (2001). Learning through computer-based concept mapping with scaffolding aid. *Journal of Computer Assisted Learning*, pages 21–33.
- Chein, M. and Mugnier, M.-L. (2008). Graph-based Knowledge Representation: Computational Foundations of Conceptual Graphs. Springer Publishing Company, Incorporated, 1 edition.
- Chen, Y., Cheng, X., and Huang, Y. (2008). A wavelet-based model to recognize high-quality topics on web forum. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, pages 343–351. IEEE.
- Coltheart, M. (2005). Word recognition processes in reading. modeling reading : the dual-route approach. In Margaret J Snowling; Charles Hulme, e., editor, *The science of reading : a handbook*. Malden, MA: Blackwell Pub.
- Cordón, L. A. and Day, J. D. (1996). Strategy use on standardized reading comprehension tests. *Journal of Educational Psychology*, 88(2):288–295.
- Cormen, T., Leiserson, C., and Stein, R. (2007). *Algoritmos*. Campus - RJ.
- Damasceno, F., Reategui, E., Schmitz, C., Harzheim, E., and Epstein, D. (2014). Supporting teleconsulting with text mining: Continuing professional development in the telehealthrs project. In *Collaboration and Technology*, volume 8658 of *Lecture Notes in Computer Science*, pages 97–104. Springer International Publishing.
- Dehaene, S. (2009). *Reading in the Brain: The New Science of How We Read*. Penguin Publishing Group.
- Dehaene, S., Jobert, A., Naccache, L., Ciuciu, P., Poline, J., Bihan, D. L., and Cohen, L. (2004). Letter binding and invariant recognition of masked words: Behavioral and neuroimaging evidence. *Psychological Science*, 15:307–313.
- Dickinson, P., Bunke, H., Dadej, A., and Kraetzl, M. (2002). Median graphs and anomalous change detection in communication networks. In *Information, Decision and Control, 2002. Final Program and Abstracts*, pages 59–64.

- Dirks, E., Spyer, G., van Lieshoul, E. C. D., and de Sonnevill, L. (2008). Prevalence of combined reading and arithmetic disabilities. *Journal of Learning Disabilities*, pages 460—473.
- Duncan, L. G. and Seymour, P. H. (2003). How do children read multisyllabic words? some preliminary observations. *Journal of Research in Reading*, 26(2):101–120.
- Ehri, Linnea C.; Wilce, L. S. (1979). The mnemonic value of orthography among begin- ning readers. *Journal of Educational Psychology*.
- Feldman, R. and Sanger, J. (2006). *Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, New York, NY, USA.
- Fisher, D., Frey, N., and D., W. (2002). Seven literary strategies that work. *Educational Leadership*, 60(3).
- Forster, K. I. and Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, 12(6):627 – 635.
- Gao, X., Murugesan, S., and Lo, B. W. N. (2005). Extraction of keyterms by simple text mining for business information retrieval. In Lau, F. C. M., Lei, H., Meng, X., and Wang, M., editors, *ICEBE*, pages 332–339. IEEE Computer Society.
- Geva, T. and Zahavi, J. (2010). Predicting intraday stock returns by integrating market data and financial news reports. In *MCIS*, page 39. AISel.
- Goldinger, S. D. and Azuma, T. (2003). Puzzle-solving science: the quixotic quest for units in speech perception. *Journal of Phonetics*, 31(3-4):305–320.
- Gowin, D. (1981). *Educating*. Cornell University Press.
- Holden, C. (1992). Study flunks science and math tests. *Science*, 258. American Association for the Advancement of Science. 541–541.
- Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. Wiley, New York, new ed edition.
- Jenner, J. (2003). A bridge to reading and writing literacy: Developing oral language skills in young children.

- Kibby, M. Y. and Hynd, G. W. (2001). Neurobiological basis of learning disabilities. In Hallahan, D. P. and (Eds.), B. K. K., editors, *Research and global perspectives in learning disabilities*, pages 25–42. Mahwah, NJ: Lawrence Erlbaum Associates.
- Korhonen, J., Linnanmäki, K., and Aunio, P. (2014). Learning difficulties, academic well-being and educational dropout: A person-centred approach. *Learning and Individual Differences*, 31(0):1 – 10.
- Krallinger, M. and Valencia, A. (2005). Text-mining and information-retrieval services for molecular biology. *Genome Biology*, 6(7):224+.
- LaBerge, D. and Samuels, S. J. (1974). Towards a theory of automatic information processing in reading. *Cognitive Psychology*, 6:292–323.
- Langa, N. R. (2014). Jogo digital para apoio ao letramento empregando técnicas de mineração de texto. Dissertação de Mestrado. Universidade Federal do Rio Grande do Sul.
- Lorenzatti, A. (2007). Sobek: uma ferramenta de mineração de textos.
- Macedo, A. L. (2010). *Rede de conceitos : uma ferramenta para contribuir com a prática pedagógica no acompanhamento da produção textual coletiva*. PhD thesis, Universidade Federal do Rio Grande do Sul.
- Macedo, A. L., Reategui, E. B., Lorenzatti, A., and Behar, P. A. (2009). Using text-mining to support the evaluation of texts produced collaboratively. In *WCCE*, volume 302 of *IFIP Advances in Information and Communication Technology*, pages 368–377. Springer.
- Macnamara, J. (1982). *Names for Things: A Study of Human Learning*. Bradford Books. MIT Press.
- Manoli, P. & Papadopoulou, M. (2012). Graphic Organizers as a Reading Strategy: Research Findings and Issues. *Creative Education*, 3, 348-356. doi: 10.4236/ce.2012.33055.
- Marcuschi, L. A.; Xavier, A. C. (2005). Hipertexto e generos digitais: novas formas de construção do sentido.
- Marzano, R., Pickering, D., and Pollock, J. (2001). *Classroom Instruction that Works: Research-based Strategies for Increasing Student Achievement*. Gale virtual

- reference library. Association for Supervision and Curriculum Development.
- Maurer, U., Brem, S., Kranz, F., Bucher, K., Benz, R., Halder, P., Steinhausen, H.-C., and Brandeis, D. (2006). Coarse neural tuning for print peaks when children learn to read. *Neuroimage*, 33(2):749–758.
- Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology*, 88(1), 64-73.
- McCandliss, B. D., Cohen, L., and Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, 7:293–299.
- McKeown, M. G., Beck, I. L., Omanson, R. C., and Perfetti, C. A. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. *Journal of Literacy Research*, 15(1):3–18.
- McKoon, G. and Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99(3):440–466.
- Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information.
- Moats, L. (1998). *Reading, spelling, and writing disabilities in the middle grades*. Learning about learning disabilities (2nd ed.), San Diego, CA: Academic Press.
- Moreira, M. (2010). Mapas Conceituais e aprendizagem significativa. Centauro.
- Nandhini, K. and Balasundaram, S. (2013). Improving readability through extractive summarization for learners with reading difficulties. *Egyptian Informatics Journal*, 14(3):195 – 204.
- Nesbit, J. C. and Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, 76(3):413–448.
- Novak, J. D. (1991). Clarify with concept maps: A tool for students and teachers alike. *The Science Teacher*, 58:45–49.
- Novak, J. D. (2002). Meaningful learning: The essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. *Science Education*, 86(4):548–571.
- Novak, J. D. and Cañas, A. J. (2006). The origins of the concept mapping tool and

- the continuing evolution of the tool. *Information Visualization*, 5(3):175–184.
- Novak, J. and Canas, A. (2007). Theoretical origins of concept maps, how to construct them, and uses in education. *Reflecting Education*, Volume 3.
- OECD (2013). Pisa 2012 results: What students know and can do. student performance in mathematics, reading and science (volume i). 1.
- Palmiero, M., Olivetti Belardinelli, M., Nardo, D., Sestieri, C., Di Matteo, R., D’Ausilio, A., and Romani, G. (2009). Mental imagery generation in different modalities activates sensory-motor areas. *Cognitive Processing*, 10(2):268–271.
- Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135.
- Pinho, I. da Costa, Epstein, D., Reategui, E. B., Correa, Y., and Polonia, E. (2013). The use of text mining to build a pedagogical agent capable of mediating synchronous online discussions in the context of foreign language learning. *2013 IEEE Frontiers in Education Conference (FIE)*, pages 393–399.
- Pressley, M. (2000). *What should comprehension instruction be the instruction of?* Hand- book of reading research, Volume 3.
- Pressley, M. and Afflerbach, P. (1995). *Verbal protocols of reading: the nature of con- structively responsive reading*. Lawrence Erlbaum Associates.
- Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., and Others (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facili- tates learning. *Educational Psychologist*.
- Rawson, K. A., Dunlosky, J., and Thiede, K. W. (2000). The rereading effect: meta- comprehension accuracy improves across reading trials. *Memory & Cognition*, 28(6):1004–1010.
- Reader, W. and Hammond, N. (1994). Computer-based tools to support learning from hypertext: concept mapping tools and beyond. *Computers and Education*, 22.
- Regina, S. and Education, S. (1992). *The adaptive dimension in core curriculum*. Saskatchewan Learning.
- Rodd, J. M., Davis, M. H., and S, I. (2005). The neural mechanisms of speech compre- hension: fmri studies of semantic ambiguity. *Cerebral Cortex*, pages

1261–1269.

- Rosenthal, J. and Ehri, L. C. (2008). The mnemonic value of orthography for vocabulary learning. *Journal of Educational Psychology*.
- Ruddell, M. R. (1997). *Teaching Content Reading and Writing*. Wiley.
- Sandak, R., Mencl, W. E., Frost, S. J., and Pugh, K. R. (2004). The neurobiological basis of skilled and impaired reading: Recent findings and new directions. *Scientific Studies of Reading*, 8(3):273–292.
- Schenker, A. (2003). *Graph-Theoretic Techniques for Web Content Mining*. PhD thesis.
- Schenker, A., Last, M., Bunke, H., and Kandel, A. (2002). A comparison of two novel algorithms for clustering web documents. In *In: Proceedings of the 2nd International Workshop on Web Document Analysis WDA (2003)*.
- Seidenberg, M. S. and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96:523–568.
- Shen, W., Shen, J., and Zheng, H. (2001). Fingerprint recognition system by use of graph matching.
- Snow, C. E., Burns, M. S., and Griffin, P. (1998). *Preventing reading difficulties in young children*. National Academy Press : U.S. Dept. of Education.
- Sridhar, D. and Vaughn, S. (2001). *Social functioning of students with learning disabilities*. Research and global perspectives in learning disabilities.
- Sticht, T. (1974). *Auding and Reading: A Developmental Model*. Human Resources Research Organization.
- Strijbos, J.-W., Martens, R. L., Prins, F. J., and Jochems, W. M. (2006). Content analysis: What are they talking about? *Computers & Education*, 46(1):29 – 48.
- Tan, A. and Nicholson, T. (1997). Flashcards revisited: Training poor readers to read words faster improves their comprehension of text. *Journal of Educational Psychology*.
- Tan, A.-h. (1999). Text mining: The state of the art and the challenges. In *In Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases*, pages 65–70.
- Tankersley, K. (2003). *The threads of reading: Strategies for literacy development*.

- Tseng, Y.-H., Lin, C.-J., and Lin, Y.-I. (2007). Text mining techniques for patent analysis. *Inf. Process. Manage.*, 43(5):1216–1247.
- Venkatesh, V., Morris, M. G., Davis, G. B., Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27(3): 425-478.
- Vieira, M. C. T. (1981). Levantamento das dificuldades de alunos do 1<sup>o</sup> ano da universidade na compreensão de materiais escritos.
- Villalon, J., Kearney, P., Calvo, R., and Reimann, P. (2008). Glosser: Enhanced feedback for student writing tasks. In *Advanced Learning Technologies, 2008. ICALT '08. Eighth IEEE International Conference on*, pages 454–458.
- Warschauer, M. (2006). *Laptops and Literacy: Learning in the Wireless Classroom*. New York: Teachers College Press.
- Wawryk-Epp, L. and Learning, S. S. (2004). *Teaching Students with Reading Difficulties and Disabilities : a Guide for Educators*. Saskatchewan Learning.
- Wei, C.-W., Hsieh, Z.-H., Chen, N.-S., and Kinshuk (2012). Construction of reading guidance mechanism on e-book reader applications for improving learners' english comprehension capabilities. In *ICALT*, pages 170–172. IEEE.
- Weinstein, L. and Siever, D. (2003). *Reading David: A Mother and Son's Journey Through the Labyrinth of Dyslexia*. Berkley Publishing Group.
- Wiskott, L., Fellous, J.-M., Kuiger, N., and von der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):775–779.

# Apêndice A



**Universidade Federal do Rio Grande do Sul**

**Programa de Pós-Graduação em Informática na Educação (PPGIE)**

## **TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO**

O pesquisador Daniel Epstein, doutorando do curso de Pós-Graduação em Informática na Educação (PPGIE) da Universidade Federal do Rio Grande do Sul (UFRGS), orientado pelo professor Dr. Eliseo Berni Reategui, convida você a participar da pesquisa intitulada **“USO DO MINERADOR DE TEXTOS SOBEK COMO FERRAMENTA DE APOIO A LEITURA E COMPREENSÃO TEXTUAL”**. A pesquisa tem por objetivo auxiliar os estudantes em atividades de leitura utilizando um programa de computador.

Os alunos e/ou responsáveis que derem seu consentimento para participação na pesquisa autorizam a coleta de dados a partir de uma atividade educacional, questionários e entrevistas.

Os dados desta pesquisa são sigilosos, não sendo mencionados os nomes dos participantes em nenhuma apresentação oral ou trabalho escrito que venha a ser publicado. A participação não oferece risco ou prejuízo ao participante. Se no decorrer da pesquisa o participante resolver interromper sua participação, poderá fazê-lo sem que isso lhe acarrete qualquer prejuízo ou constrangimento.

O pesquisador compromete-se a esclarecer quaisquer dúvidas ou questionamentos que eventualmente os participantes venham a ter no momento da pesquisa ou posteriormente através do e-mail [daepstein@hotmail.com](mailto:daepstein@hotmail.com).

Após ter sido devidamente informado(a) de todos os aspectos desta pesquisa:

Eu, \_\_\_\_\_, concordo em participar desta pesquisa.

(Nome do aluno)

\_\_\_\_\_

Assinatura do/da aluno

\_\_\_\_\_

Assinatura do (a) responsável

Porto Alegre, \_\_\_\_ de \_\_\_\_\_ de 2015.

# Apêndice B

## Termo de consentimento da instituição de ensino

Eu, Daniel Epstein, doutorando do Programa de Pós-Graduação em Informática na Educação da Universidade Federal do Rio Grande do Sul, responsável pelo projeto intitulado **Uso do Minerador de Textos Sobek como Ferramenta de apoio à Leitura e Compreensão Textual**, venho pelo presente termo solicitar sua autorização para realizar este projeto de pesquisa na Escola Municipal Ludovino Fanton.

O projeto tem como objetivo principal investigar como uma ferramenta computacional capaz de extrair conceitos de textos pode contribuir com o processo de leitura e compreensão textual. A participação dos alunos consistirá em realizar atividades de leitura e escrita utilizando a ferramenta computacional. A principal ferramenta a ser utilizada pode ser encontrada no endereço de internet: <http://sobek.ufrgs.br>. Os dados coletados durante as atividades consistirão nas produções escritas dos alunos, questionários e entrevistas com os alunos e professores responsáveis com relação ao uso da ferramenta proposta no apoio à leitura.

Quaisquer esclarecimentos sobre o desenvolvimento do projeto de pesquisa podem ser solicitados, sendo possível retirar sua autorização para o prosseguimento das atividades na escola em qualquer momento.

Os dados obtidos nesta pesquisa serão utilizados futuramente para a publicação de artigos científicos. Não serão veiculadas imagens dos alunos nem das instalações da escola. Assumimos a total responsabilidade de não publicar qualquer dado que comprometa o sigilo da participação dos integrantes de sua instituição, como nome, endereço e outras informações pessoais. A participação será voluntária e poderá ser interrompida pelos estudantes a qualquer momento, sem que haja nenhum prejuízo.

### Autorização Institucional

Eu \_\_\_\_\_ (nome legível) responsável pela **Escola Municipal Ludovino Fanton**, declaro que fui informado dos objetivos da pesquisa acima e concordo em autorizar a execução da

mesma nesta instituição. Caso necessário, a qualquer momento poderemos revogar esta autorização.

---

Pesquisador

---

Responsável pela instituição

Data

# Apêndice C

## Instruções para a resposta do questionário do aluno

As questões de 1 – 3 possuem as seguintes possibilidades de respostas (respostas no padrão escala Likert de 7 pontos):

- Completamente satisfeito
- Bastante satisfeito
- Um pouco satisfeito
- Nem satisfeito nem insatisfeito
- Um pouco insatisfeito
- Bastante insatisfeito
- Completamente insatisfeito

As questões 4 e 5 possuem as seguintes possibilidades de respostas (respostas no padrão escala Likert de 5 pontos):

- Muito fácil
- Fácil
- Neutro
- Difícil
- Muito difícil

As questões de 6 – 8 possuem as seguintes possibilidades de respostas:

- 0-1
- 2-3
- 4-5
- 5-6
- 6+

As questões de 12 – 14 possuem possibilidades de resposta livre.

## QUESTIONÁRIO DE AVALIAÇÃO DO ESTUDANTE

Este questionário tem como objetivo avaliar a atividade desenvolvida com o auxílio da ferramenta Sobek. Esta ferramenta faz parte de um projeto de doutorado orientado pelo professor doutor Eliseo Reategui vinculado ao Programa de Pós-Graduação em Informática na Educação da Universidade Federal do Rio Grande do Sul. A fim de obtermos subsídios que nos permitam avaliar a ferramenta solicitamos que todas as questões sejam respondidas. Sua contribuição é muito importante para que possamos prosseguir buscando sempre o aperfeiçoamento do projeto. Sua participação na pesquisa é totalmente voluntária e as informações coletadas serão totalmente confidenciais e não serão utilizados de nenhuma forma para avaliar seu desempenho. Você será esclarecido sobre a pesquisa em qualquer aspecto que desejar é livre para recusar se a participar retirar seu consentimento ou interromper a participação a qualquer momento. A recusa em participar não irá acarretar qualquer penalidade ou perda de benefício. Para confirmar sua participação voluntária na pesquisa selecione a opção desejada preencha seu nome e responda às questões a seguir.

Você gostaria de participar da pesquisa? **(Sim/Não)**

*Nome Completo*

*Ano/Serie*

*Turma*

*Instituição*

1. Qual o seu nível de satisfação em relação atividade que acabou de participar?
2. Qual o seu nível de satisfação em relação à ferramenta Sobek?
3. Qual o seu nível de satisfação em relação ao grafo fornecido pelo Sobek para o texto da atividade?
4. Qual o grau de dificuldade para a utilização da ferramenta Sobek?
5. Qual o grau de dificuldade para a interação com o grafo?
6. Quantos termos você adicionou ao grafo?
7. Quantos termos você removeu do grafo?
8. Quantas relações você editou no grafo?
9. Você precisou da ajuda do professor durante a atividade? **(Sim/ Não sei/ Não)**
10. Você considera que o Sobek o auxiliou a compreender o texto? **(Sim/ Não sei/ Não)**

11. Você utilizaria o Sobek fora da sala de aula? (**Sim/ Não sei/ Não**)
12. Comente os aspectos que você considera POSITIVO sobre a atividade ou sobre a ferramenta Sobek.
13. Comente os aspectos que você considera NEGATIVO sobre a atividade ou sobre a ferramenta Sobek
14. Você tem alguma outra sugestão ou crítica sobre a atividade ou sobre a ferramenta Sobek?

# Apêndice D

## Questionário com a professora Gabriela, professora concursada da escola municipal Lidivino Fanton

Dados do professor:

- Nome
- Disciplina ministrada
- Instituição
- Idade
- Tempo de formação

### GUIA DE QUESTÕES PARA O PROFESSOR

1. Qual a opinião do professor acerca da atividade realizada?
2. A ferramenta é simples de utilizar?
3. O professor usaria a ferramenta durante suas aulas?
4. Quais as principais dificuldades em preparar a atividade?
5. Qual a relação do professor com recursos tecnológicos pedagógicos?
6. Qual a avaliação do professor acerca do desempenho dos estudantes durante a atividade?
7. Toda a atividade transcorreu em sala de aula?
8. Algum aluno solicitou auxílio para o uso da ferramenta ou desenvolvimento da atividade?
9. Qual a avaliação geral do professor acerca dos aspectos pedagógicos da ferramenta?

# Apêndice E



**Universidade Federal do Rio Grande do Sul**

**Programa de Pós-Graduação em Informática na Educação (PPGIE)**

## **TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO**

O pesquisador Daniel Epstein, doutorando do curso de Pós-Graduação em Informática na Educação (PPGIE) da Universidade Federal do Rio Grande do Sul (UFRGS), orientado pelo professor Dr. Eliseo Berni Reategui, convida você a participar da pesquisa intitulada **“USO DO MINERADOR DE TEXTOS SOBEK COMO FERRAMENTA DE APOIO A LEITURA E COMPREENSÃO TEXTUAL”**. A pesquisa tem por objetivo auxiliar os estudantes em atividades de leitura utilizando um programa de computador.

Os alunos e/ou responsáveis que derem seu consentimento para participação na pesquisa autorizam a coleta de dados a partir de uma atividade educacional, questionários e entrevistas.

Os dados desta pesquisa são sigilosos, não sendo mencionados os nomes dos participantes em nenhuma apresentação oral ou trabalho escrito que venha a ser publicado. A participação não oferece risco ou prejuízo ao participante. Se no decorrer da pesquisa o participante resolver interromper sua participação, poderá fazê-lo sem que isso lhe acarrete qualquer prejuízo ou constrangimento.

O pesquisador compromete-se a esclarecer quaisquer dúvidas ou questionamentos que eventualmente os participantes venham a ter no momento da pesquisa ou posteriormente através do e-mail [daepstein@hotmail.com](mailto:daepstein@hotmail.com).

Após ter sido devidamente informado(a) de todos os aspectos desta pesquisa:

Eu, \_\_\_\_\_, concordo em participar desta pesquisa.

(Nome do aluno)

\_\_\_\_\_

Assinatura do/da aluno

\_\_\_\_\_

Assinatura do (a) responsável

Porto Alegre, \_\_\_\_ de \_\_\_\_\_ de 2017.

# Apêndice F



**Universidade Federal do Rio Grande do Sul**

**Programa de Pós-Graduação em Informática na Educação (PPGIE)**

## **Termo de consentimento da instituição de ensino**

Eu, Daniel Epstein, doutorando do Programa de Pós-Graduação em Informática na Educação da Universidade Federal do Rio Grande do Sul, responsável pelo projeto intitulado **Uso do Minerador de Textos Sobek como Ferramenta de apoio à Leitura e Compreensão Textual**, venho pelo presente termo solicitar sua autorização para realizar este projeto de pesquisa no **Colégio Murialdo**.

O projeto tem como objetivo principal investigar como uma ferramenta computacional capaz de extrair conceitos de textos pode contribuir com o processo de leitura e compreensão textual. A participação dos alunos consistirá em realizar atividades de leitura e escrita utilizando a ferramenta computacional. A principal ferramenta a ser utilizada pode ser encontrada no endereço de internet: <http://sobek.ufrgs.br>. Os dados coletados durante as atividades consistirão nas produções escritas dos alunos, questionários e entrevistas com os alunos e professores responsáveis com relação ao uso da ferramenta proposta no apoio à leitura.

Quaisquer esclarecimentos sobre o desenvolvimento do projeto de pesquisa podem ser solicitados, sendo possível retirar sua autorização para o prosseguimento das atividades na escola em qualquer momento.

Os dados obtidos nesta pesquisa serão utilizados futuramente para a publicação de artigos científicos. Não serão veiculadas imagens dos alunos nem das instalações da escola. Assumimos a total responsabilidade de não publicar qualquer dado que comprometa o sigilo da participação dos integrantes de sua instituição, como nome, endereço e outras informações pessoais. A participação será voluntária e poderá ser interrompida pelos estudantes a qualquer momento, sem que haja nenhum prejuízo.

### **Autorização Institucional**

Eu \_\_\_\_\_ (nome legível) responsável pelo Colégio Murialdo, declaro que fui informado dos objetivos da pesquisa acima e concordo em autorizar a execução da mesma nesta instituição. Caso necessário, a qualquer momento poderemos revogar esta autorização.

\_\_\_\_\_

Pesquisador

\_\_\_\_\_

Responsável pela instituição

Data

# Apêndice G

## Textos e questionários utilizados na primeira intervenção realizada no Colégio Murialdo junto aos alunos de Ciclo 1

### As Amazônias

Esse tapete de florestas com rios azuis que os astronautas viram é a Amazônia. Ela cobre mais da metade do território brasileiro. Quem viaja pela região, não cansa de admirar as belezas da maior floresta tropical do mundo. No início era assim: água e céu.

É mata que não tem mais fim. Mata contínua, com árvores muito altas, cortada pelo Amazonas, o maior rio do planeta. São mais de mil rios desaguando no Amazonas. É água que não acaba mais.

Fonte: SALDANHA, P. As Amazônias. Rio de Janeiro: Ediouro, 1995

### Questionário

**1) No texto, o uso da expressão “água que não acaba mais” revela**

- (A) admiração pelo tamanho do rio.
- (B) ambição pela riqueza da região.
- (C) medo da violência das águas.
- (D) surpresa pela localização do rio.

**2) O texto trata**

- (A) da importância econômica do rio Amazonas.

- (B) das características da região Amazônica.
- (C) de um roteiro turístico da região do Amazonas.
- (D) do levantamento da vegetação amazônica.

**3) A frase que contém uma opinião é**

- (A) “cobre mais da metade do território brasileiro”.
- (B) “não cansa de admirar as belezas da maior floresta”.
- (C) “...maior floresta tropical do mundo”.
- (D) “é Mata contínua [...] cortada pelo Amazonas”.

### **Feias, sujas e imbatíveis**

As baratas estão na Terra há mais de 200 milhões de anos, sobrevivem tanto no deserto como nos polos e podem ficar até 30 dias sem comer. Vai encarar?

Férias, sol e praia são alguns dos bons motivos para comemorar a chegada do verão e achar que essa é a melhor estação do ano. E realmente seria, se não fosse por um único detalhe: as baratas. Assim como nós, elas também ficam bem animadas com o calor. Aproveitam a aceleração de seus processos bioquímicos para se reproduzirem mais rápido e, claro, para passearem livremente por todos os cômodos de nossas casas.

Nessa época do ano, as chances de dar de cara com a visitante indesejada, ao acordar durante a noite para beber água ou ir ao banheiro, são três vezes maiores.

Fonte: Revista Galileu. Rio de Janeiro: Globo, Nº 151, Fev. 2004, p.26.

## Questionário

1) No trecho “Vai encarar?” (linha 2), o ponto de interrogação tem o efeito de

- a) Apresentar
- b) Avisar
- c) Desafiar
- d) Questionar

2) A expressão “Vai encarar?” (linha 2), é marca de linguagem

- a) Científica
- b) Formal
- c) Informal
- d) Regional

---

### **Textos e questionários utilizados na primeira intervenção realizada no Colégio Murialdo junto aos alunos de Ciclo 3**

#### **IMIM 100**

Yoshiko era apenas uma criança quando saiu do Japão, sua terra-natal. Depois de 52 dias de viagem no navio Kasato Maru, finalmente ela desembarcou no Brasil com seus pais e um irmão em 18 de junho de 1908. No mesmo navio, havia outras 164 famílias japonesas, todas em busca de trabalho e melhores condições de vida. Por aqui, Yoshiko encontrou muitas coisas diferentes. Estranhou a comida, a língua, as roupas, o clima... Mas o jeito era encarar o trabalho nas lavouras de café e ajuntar dinheiro logo para

regressar ao Japão. Assim, Yoshiko e sua família se instalaram no interior de São Paulo. A vida não era fácil e o salário também não era dos melhores. Mesmo assim, a cada ano, mais e mais japoneses cruzavam o oceano em direção ao Brasil.

O tempo passou. Quando Yoshiko e sua família finalmente conseguiram juntar um bom dinheiro, a Segunda Guerra Mundial estourou. Aí, tudo ficou mais difícil. O jeito foi continuar no Brasil por mais uns anos. Só que, com o fim da Guerra, em 1945, já não fazia mais sentido voltar para o Japão. Yoshiko conheceu um outro imigrante japonês e se casou, formando uma família no Brasil.

Nos anos 60, os filhos de Yoshiko decidiram se mudar para a cidade grande para estudar. Como outros filhos de imigrantes também tomaram essa decisão, cidades como São Paulo ficaram lotadas de japoneses, principalmente no bairro da Liberdade.

Fonte: Revista Nosso Amiguinho. Pesquisa. Texto: Fernando Torres. Junho de 2008. p.17.

### **Questionário**

#### **1) O que mudou nos planos de Yoshiko com a Segunda Guerra Mundial?**

- a) Muitos imigrantes japoneses chegaram ao Brasil.
- b) Yoshiko e sua família permaneceram no Brasil por mais alguns anos.
- c) Yoshiko e sua família tiveram que trabalhar nas lavouras de café.
- d) Onavio Kasato Maru iniciou uma viagem de 52 dias.

#### **2) O que o texto fala sobre o bairro da Liberdade?**

- a) Era o bairro onde Yoshiko morava quando chegou no Brasil.
- b) Muitos imigrantes foram para lá durante a Segunda Guerra Mundial.
- c) Foi um bairro criado nos anos 60.

d) Bairro que possui muitos moradores japoneses.

## **Vínculos, as equações da matemática da vida**

Quando você forma um vínculo com alguém, forma uma aliança. Não é à toa que o uso de alianças é um dos símbolos mais antigos e universais do casamento. O círculo dá a noção de ligação, de fluxo, de continuidade. Quando se forma um vínculo, a energia flui. E o vínculo só se mantém vivo se essa energia continuar fluindo. Essa é a ideia de mutualidade, de troca. Nessa caminhada da vida, ora andamos de mãos dadas, em sintonia, deixando a energia fluir, ora nos distanciamos. Desvios sempre existem. Podemos nos perder em um deles e nos reencontrar logo adiante. A busca é permanente. O que não se pode é ficar constantemente fora de sintonia.

Antigamente, dizia-se que as pessoas procuravam se completar através do outro, buscando sua metade no mundo. A equação era:  $1/2 + 1/2 = 1$ . "Para eu ser feliz para sempre na vida, tenho que ser a metade do outro." Naquela loteria do casamento, tirar a sorte grande era achar a sua cara-metade. Com o passar do tempo, as pessoas foram desenvolvendo um sentido de individualização maior e a equação mudou. Ficou:  $1 + 1 = 1$ . "Eu tenho que ser eu, uma pessoa inteira, com todas as minhas qualidades, meus defeitos, minhas limitações. Vou formar uma unidade com meu companheiro, que também é um ser inteiro." Mas depois que esses dois seres inteiros se encontravam, era comum fundirem-se, ficarem grudados num casamento fechado, tradicional. Anulavam-se mutuamente.

Com a revolução sexual e os movimentos de libertação feminina, o processo de individuação que vinha acontecendo se radicalizou. E a equação mudou de novo:  $1 + 1 = 1 + 1$ . Era o "cada um na sua". "Eu tenho que resolver os meus problemas, cuidar da minha própria vida. Você deve fazer o mesmo. Na minha independência total e autossuficiência absoluta, caso com você, que também é assim." Em nome dessa independência, no entanto, faltou sintonia, cumplicidade e compromisso afetivo.

É a segunda crise do casamento que acompanhamos nas décadas de 70 e 80.

Atualmente, após todas essas experiências, eu sinto as pessoas procurando outro tipo de equação:  $1 + 1 = 3$ . Para a aritmética ela pode não ter lógica, mas faz sentido do ponto de vista emocional e existencial. Existem você, eu e a nossa relação. O vínculo entre nós é algo diferente de uma simples somatória de nós dois. Nessa proposta de casamento, o que é meu é meu, o que é seu é seu e o que é nosso é nosso. Talvez aí esteja a grande mágica que hoje buscamos, a de preservar a individualidade sem destruir o vínculo afetivo. Tenho que preservar o meu eu, meu processo de descoberta, realização e crescimento, sem destruir a relação. Por outro lado, tenho que preservar o vínculo sem destruir a individualidade, sem me anular.

Acho que assim talvez possamos chegar ao ano 2000 um pouco menos divididos entre a sede de expressão individual e a fome de amor e de partilhar a vida. Um pouco mais inteiros e felizes. Para isso, temos que compartilhar com nossos companheiros de uma verdadeira intimidade. Ser íntimo é ser próximo, é estar estreitamente ligado por laços de afeição e confiança.

(MATARAZZO, Maria Helena. Amar é preciso. 22. ed. São Paulo: Editora Gente, 1992. p. 19-21)

### **Questionário**

**1) O texto trata PRINCIPALMENTE:**

- (A) da exatidão da matemática da vida.
- (B) dos movimentos de libertação feminina.
- (C) da loteria do sucesso no casamento.
- (D) do casamento no passado e no presente.

**2) No texto, no casamento, atualmente, defende-se a ideia de que:**

- (A) a felicidade está na somatória do casal.
- (B) a unidade é igual à soma das partes.
- (C) o ideal é preservar o “eu” e o vínculo afetivo.
- (D) o melhor é cada um cuidar de sua própria vida.

**3) Segundo o texto, com a revolução sexual e os movimentos de libertação feminina:**

- (A) as mulheres estavam livres para buscar uma verdadeira intimidade.
- (B) os vínculos afetivos foram preservados.
- (C) todos quiseram partilhar suas vidas.
- (D) as pessoas ficaram mais independentes e autossuficientes.

**4) De acordo com o texto, o que acontecia quando as pessoas buscavam sua metade pelo mundo?**

- (A) elas tentavam se completar através do parceiro.
- (B) aprendiam sobre aritmética e lógica.
- (C) criavam laços de confiança e afeição.
- (D) passavam por uma fase de descobertas, realizações e crescimento.

**5) Como a autora explica a equação:  $1 + 1 = 1$ ?**

- (A) uma maneira de relacionar a matemática à nossa vida.

(B) uma preocupação com os vínculos afetivos.

(C) uma prova de que o casamento mudou.

(D) as pessoas eram inteiras, mas formavam unidades com seus parceiros.

# Apêndice H

## Textos e questionários utilizados na segunda intervenção realizada no Colégio Murialdo junto aos alunos de Ciclo 1

### O dono da bola – Ruth Rocha

O nosso time estava cheio de amigos. O que nós não tínhamos era a bola de futebol. Só bola de meia, mas não é a mesma coisa.

Bom mesmo é bola de couro, como a do Caloca.

Mas, toda vez que nós íamos jogar com Caloca, acontecia a mesma coisa. E era só o juiz marcar qualquer falta do Caloca que ele gritava logo:

– Assim eu não jogo mais! Dá aqui a minha bola!

– Ah, Caloca, não vá embora, tenha espírito esportivo, jogo é jogo...

– Espírito esportivo, nada! – berrava Caloca. – E não me chame de Caloca, meu nome é Carlos Alberto!

E assim, Carlos Alberto acabava com tudo que era jogo.

A coisa começou a complicar mesmo, quando resolvemos entrar no campeonato do nosso bairro. Nós precisávamos treinar com bola de verdade para não estranhar na hora do jogo.

Mas os treinos nunca chegavam ao fim. Carlos Alberto estava sempre procurando encrenca:

– Se o Beto jogar de centroavante, eu não jogo!

– Se eu não for o capitão do time, vou embora!

– Se o treino for muito cedo, eu não trago a bola!

E quando não se fazia o que ele queria, já sabe, levava a bola embora e adeus, treino.

Catapimba, que era o secretário do clube, resolveu fazer uma reunião:

– Esta reunião é para resolver o caso do Carlos Alberto. Cada vez que ele se zanga, carrega a bola e acaba com o treino.

Carlos Alberto pulou, vermelhinho de raiva:

– A bola é minha, eu carrego quantas vezes eu quiser!

– Pois é isso mesmo! – disse o Beto, zangado. – É por isso que nós não vamos ganhar campeonato nenhum!

– Pois, azar de vocês, eu não jogo mais nessa droga de time, que nem bola tem.

E Caloca saiu pisando duro, com a bola debaixo do braço.

Aí, Carlos Alberto resolveu jogar bola sozinho. Nós passávamos pela casa dele e víamos. Ele batia bola com a parede. Acho que a parede era o único amigo que ele tinha. Mas eu acho que jogar com a parede não deve ser muito divertido.

Porque, depois de três dias, o Carlos Alberto não aguentou mais. Apareceu lá no campinho.

– Se vocês me deixarem jogar, eu empresto a minha bola.

Carlos Alberto estava outro. Jogava direitinho e não criava caso com ninguém.

E, quando nós ganhamos o jogo final do campeonato, todo mundo se abraçou gritando:

– Viva o Estrela-d’Alva Futebol Clube!

– Viva!

– Viva o Catapimba!

– Viva!

– Viva o Carlos Alberto!

– Viva!

Então o Carlos Alberto gritou:

– Ei, pessoal, não me chamem de Carlos Alberto! Podem me chamar de Caloca!

## **Questionário**

### **1 - No começo da história:**

- a) O time não jogava futebol
- b) O time tinha apenas uma bola de couro
- c) Não havia time porque não tinham amigos
- d) O time tinha só uma bola de meia

### **2 - Calota sempre parava o jogo porque:**

- a) Os jogadores eram desleais com ele
- b) Ele sofria muitas faltas durante o jogo
- c) Ele não gostava dos outros jogadores
- d) Ele queria levar vantagem por ser o dono da bola.

### **3 - O time ganhou a final do campeonato porque:**

- a) Carlos Alberto era muito bom jogador
- b) Eles compraram uma bola para treinar
- c) Caloca não gostou de ficar sozinho e decidiu emprestar a bola para jogar com os amigos
- d) O secretário Catapimba ajeitou o time durante uma reunião

## **A velha contrabandista – Sérgio Porto**

Diz que era uma velhinha que sabia andar de lambreta. Todo dia ela passava pela fronteira montada na lambreta, com um bruto saco atrás da lambreta. O pessoal da Alfândega – tudo malandro velho – começou a desconfiar da velhinha.

Um dia, quando ela vinha na lambreta com o saco atrás, o fiscal da Alfândega mandou ela parar. A velhinha parou e então o fiscal perguntou assim pra ela:

– Escuta aqui, vovozinha, a senhora passa por aqui todo dia, com esse saco aí atrás. Que diabo a senhora leva nesse saco?

A velhinha sorriu com os poucos dentes que lhe restavam e mais os outros, que ela adquirira no odontólogo e respondeu:

– É areia!

Aí quem sorriu foi o fiscal. Achou que não era areia nenhuma e mandou a velhinha saltar da lambreta para examinar o saco. A velhinha saltou, o fiscal esvaziou o saco e dentro só tinha areia. Muito encabulado, ordenou à velhinha que fosse em frente. Ela montou na lambreta e foi embora, com o saco de areia atrás.

Mas o fiscal ficou desconfiado ainda. Talvez a velhinha passasse um dia com areia e no outro com muamba, dentro daquele maldito saco. No dia seguinte, quando ela passou na lambreta com o saco atrás, o fiscal mandou parar outra vez. Perguntou o que é que ela levava no saco e ela respondeu que era areia, uai! O fiscal examinou e era mesmo. Durante um mês seguido o fiscal interceptou a velhinha e, todas as vezes, o que ela levava no saco era areia.

Diz que foi aí que o fiscal se chateou:

– Olha, vovozinha, eu sou fiscal de alfândega com 40 anos de serviço. Manjo essa coisa de contrabando pra burro. Ninguém me tira da cabeça que a senhora é contrabandista.

– Mas no saco só tem areia! – insistiu a velhinha. E já ia tocar a lambreta, quando o fiscal propôs:

– Eu prometo à senhora que deixo a senhora passar. Não dou parte, não apreendo, não conto nada a ninguém, mas a senhora vai me dizer: qual é o contrabando que a senhora está passando por aqui todos os dias?

– O senhor promete que não “espaia” ? – quis saber a velhinha.

– Juro – respondeu o fiscal.

– É lambreta.

## Questionário

### **1 - O fiscal tinha certeza que a velhinha estava com contrabando porque:**

- a) Pessoas velhas não andam de lambreta
- b) A velhinha estava carregando um saco muito pesado
- c) Ele era um fiscal "malandro velho"
- d) A velhinha passava todos os dias pela fronteira com um saco na moto

### **2 - Quando o fiscal pediu para olhar o saco que estava atrás da lambreta, a velhinha sorriu porque:**

- a) Não tinha contrabando no saco
- b) Quis mostrar os dentes que ainda restavam e os dentes novos adquiridos no odontólogo
- c) Era mais velha que o velho fiscal
- d) Achou que o fiscal tinha descoberto o contrabando dela

### **3 - No final da história**

- a) A velhinha não tinha contrabando

- b) O fiscal prendeu a velhinha
- c) O fiscal achou que a velhinha não tinha nenhum contrabando
- d) A velhinha confiou no fiscal

# Apêndice I

## Textos e questionários utilizados na segunda intervenção realizada no Colégio Murialdo junto aos alunos de Ciclo 3

### Defenestração - Luis Fernando Verissimo

Certas palavras têm o significado errado. Falácia, por exemplo, devia ser o nome de alguma coisa vagamente vegetal. As pessoas deveriam criar falácias em todas as suas variedades. A Falácia Amazônica. A misteriosa Falácia Negra. Hermeneuta deveria ser o membro de uma seita de andarilhos herméticos. Onde eles chegassem, tudo se complicaria.

– Os hermeneutas estão chegando!

– Ih, agora é que ninguém vai entender mais nada...

Os hermeneutas ocupariam a cidade e paralisariam todas as atividades produtivas com seus enigmas e frases ambíguas. Ao se retirarem deixariam a população prostrada pela confusão. Levaria semanas até que as coisas recuperassem o seu sentido óbvio. Antes disso, tudo pareceria ter um sentido oculto. (...)

Traquinagem devia ser uma peça mecânica.

– Vamos ter que trocar a traquinagem. E o vetor está gasto.

Mas nenhuma palavra me fascinava tanto quanto defenestração. A princípio foi o fascínio da ignorância. Eu não sabia o seu significado, nunca lembrava de procurar no dicionário e imaginava coisas. Defenestrar devia ser um ato exótico praticado por poucas pessoas. Tinha até um certo tom lúbrico.

Galanteadores de calçada deviam sussurrar no ouvido das mulheres:

– Defenestras?

A resposta seria um tapa na cara. Mas algumas... Ah, algumas defenestravam.

Também podia ser algo contra pragas e insetos. As pessoas talvez mandassem defenestrar a casa. Haveria, assim, defenestradores profissionais. Ou quem sabe seria uma daquelas misteriosas palavras que encerravam os documentos formais? “Nestes termos, pede defenestração...” Era uma palavra cheia de implicações. Devo até tê-la usado uma ou outra vez, como em:

– Aquele é um defenestrado.

Dando a entender que era uma pessoa, assim, como dizer? Defenestrada. Mesmo errada, era a palavra exata. Um dia, finalmente, procurei no dicionário. E aí está o Aurelião que não me deixa mentir. “Defenestração” vem do francês “defenestration”. Substantivo feminino. Ato de atirar alguém ou algo pela janela.

Ato de atirar alguém ou algo pela janela! Acabou a minha ignorância, mas não a minha fascinação. Um ato como este só tem nome próprio e lugar nos dicionários por alguma razão muito forte. Afinal, não existe, que eu saiba, nenhuma palavra para o ato de atirar alguém ou algo pela porta, ou escada abaixo. Por que, então, defenestração?

Talvez fosse um hábito francês que caiu em desuso.(...)

Quem entre nós nunca sentiu a compulsão de atirar alguém ou algo pela janela? A basculante foi inventada para desencorajar a defenestração. Toda a arquitetura moderna, com suas paredes externas de vidro reforçado e sem aberturas, pode ser uma reação inconsciente a esta volúpia humana, nunca totalmente dominada. Na lua-de-mel, numa suíte matrimonial no 17º andar.

Em uma ocasião, uma multidão cerca o homem que acaba de cair na calçada. Entre gemidos, ele aponta para cima e balbucia:

– Fui defenestrado...

Alguém comenta:

– Coitado. E depois ainda atiraram ele pela janela?

Agora mesmo me deu uma estranha compulsão de arrancar o papel da máquina, amassá-lo e defenestrar esta crônica. Se ela sair, é porque resisti.

### **Questionário**

#### **1. De acordo com esse texto, o que é defenestração?**

- a) Dedetizar insetos pelas ruas.
- b) Fazer solicitação ao juiz.
- c) Galantear alguém nas calçadas.
- d) Atirar algo ou alguém pela janela.
- e) Uma peça mecânica.

#### **2. No decorrer do texto, o narrador imaginava possíveis significados para defenestração. Pela ordem, poderíamos dizer que ele atribuía à palavra os seguintes sentidos:**

- a) conduta não convencional, dedetização, deferimento.
- b) prática ilegal, arrumação, requerimento.
- c) xingamento, cuidados domésticos, documento formal.
- d) indiscrição, arrumação, providências.

e) conduta imprópria, consertos, aprovação.

**3. Depois de descobrir o real significado de defenestração, o narrador continua fascinado pela palavra porque:**

a) o significado dela remete a um ato pouco comum, e ele fica imaginando as razões da existência de tal palavra.

b) ele não havia pensado na possibilidade do significado real.

c) ele não vê utilidade na palavra.

d) ele não se mostra favorável a estrangeirismos.

e) o significado da palavra remete a uma ação que não praticamos em nossa cultura.

### **Química da Digestão – Autor desconhecido**

Para viver, entre outras coisas, precisamos de energia. Como não podemos tirar energia da luz do sol para viver, como os vegetais, essa energia usada pelo nosso organismo vem das reações químicas que acontecem nas nossas células.

Podemos nos comparar a uma fábrica que funciona 24 horas por dia. Vivemos fazendo e refazendo os materiais de nossas células. Quando andamos, cantamos, pensamos, trabalhamos ou brincamos, estamos consumindo energia química gerada pelo nosso próprio organismo. E o nosso combustível vem dos alimentos que comemos.

No motor do carro, por exemplo, a gasolina ou o álcool misturam-se com o ar, produzindo uma combustão, que é uma reação química entre o combustível e o oxigênio do ar. Do mesmo modo, nas células do nosso organismo, os alimentos reagem com o oxigênio para produzir energia. No nosso corpo, os organismos são transformados nos seus componentes mais simples, equivalentes à gasolina ou ao álcool, e, portanto, mais fáceis de queimar. O processo se faz através de um grande número de reações químicas

que começam a se produzir na boca, seguem no estômago e acabam nos intestinos. As substâncias presentes nesses alimentos são decompostas pelos fermentos digestivos e se transformam em substâncias orgânicas mais simples. Daí esses componentes são transportados pelo sangue até as células. Tudo isso também consome energia.

A energia necessária para todas essas transformações é produzida pela reação química entre esses componentes mais simples, que são o nosso combustível e o oxigênio do ar. Essa é uma verdadeira combustão, mas uma combustão sem chamas, que se faz dentro de pequenas formações que existem nas células, as mitocôndrias, que são nossas verdadeiras usinas de energia.

### **Questionário**

**1 – O texto afirma que o nosso corpo pode ser comparado a uma fábrica porque:**

- a) reage quimicamente pela combustão
- b) move-se a base de gasolina ou álcool
- c) produz energia a partir dos alimentos
- d) utiliza oxigênio como combustível
- e) Funciona 22 horas por dia

**2 – “Tudo isso também consome energia” (3º parágrafo ). No trecho, a expressão em destaque se refere a:**

- a) Fermentos digestivos
- b) combustíveis
- c) reações químicas
- d) usinas de energia
- e) energia

**3 – Depois de processadas pelos fermentos digestivos, as substâncias são levadas para:**

- a) a boca
- b) as células
- c) o estômago
- d) os intestinos
- e) o esôfago

**4 – O texto trata:**

- a) da constituição do aparelho digestivo
- b) da digestão como fonte de energia
- c) dos cuidados para uma boa alimentação
- d) dos elementos que compõem o corpo humano
- e) do processo da degustação

# Apêndice J

**Questionário de avaliação da atividade realizada e do minerador de textos Sobek,  
fornecido aos estudantes do Ciclo 3 do estudo final desta tese.**

**1) Qual o seu nível de satisfação em relação as atividades realizadas?**

*Detalhamento da questão: Quanto você gostou de participar das atividades.*

**Opções:**

Gostei bastante

Gostei

Nem gostei nem achei ruim

Achei ruim

Achei bastante ruim

**2) Qual o seu nível de satisfação em relação à ferramenta Sobek?**

*Detalhamento da questão: O que você achou da ferramenta de mineração que utilizou*

**Opções:**

Gostei bastante

Gostei

Nem gostei nem achei ruim

Achei ruim

Achei bastante ruim

**3) Você considera que o grafo representa o texto?**

*Detalhamento da questão: Os grafos criados pelo Sobek mostravam o que o texto queria dizer?*

**Opções:**

Representou muito bem

Representou bem

Mais ou menos

Representou mal

Representou muito mal

**4) Qual o grau de dificuldade para utilizar a ferramenta Sobek?**

*Detalhamento da questão: Foi muito difícil utilizar o minerador?*

**Opções:**

Muito fácil

Fácil

Nem fácil nem difícil

Difícil

Muito difícil

**5) Nas atividade de leitura, você considera que o Sobek ajudou a entender o texto?**

**Opções:**

Sim

Não sei

Não

**6) Você utilizaria o Sobek fora da sala de aula?**

**Opções:**

Sim

Não sei

Não

**7) Quais os pontos bons do Sobek?**

*(Resposta dissertativa)*

**8) Quais os pontos ruins do Sobek?**

*Detalhamento da questão: Seja sincero, isso vai nos ajudar a melhorar o Sobek. Não tem problema em dizer coisas ruins dele.*

*(Resposta dissertativa)*

**9) Como você acha que o Sobek pode melhorar?**

*Detalhamento da questão: O que você gostaria de ver no Sobek ou de retirar dele? Seja sincero, isso vai nos ajudar a melhorar o Sobek. Não tem problema em dizer coisas ruins dele.*

*(Resposta dissertativa)*

**10) O que você achou das atividades?**

*Detalhamento da questão: Seja sincero, isso vai nos ajudar a melhorar o Sobek. Não tem problema em dizer coisas ruins dele.*

*(Resposta dissertativa)*