

Uso de mineração de textos no apoio à compreensão textual

Daniel Epstein , PGIE/UFRGS, daepstein@gmail.com
Eliseo Reategui, PGIE/UFRGS, eliseoreategui@gmail.com

Resumo

Ler e escrever uma pequena frase já foi suficiente para que uma pessoa seja considerada alfabetizada. Atualmente, alfabetização é um processo mais complexo que simplesmente ser capaz de ler e escrever frases esparsas; também é importante entender o que está sendo lido e escrito, além de poder utilizar essa informação. Essa nova definição de alfabetização nem sempre é obtida através das abordagens tradicionais de ensino. A falta de letramento afeta não apenas o rendimento escolar dos alunos, mas também suas relações pessoais e profissionais. Neste trabalho é proposto o uso de um minerador de textos chamado Sobek como ferramenta de apoio à compreensão textual. O minerador identifica os conceitos mais relevantes de um texto, as relações entre esses conceitos e apresenta os resultados em uma visualização de grafo. Um estudo acerca da capacidade de extração de informações do Sobek é apresentado.

Palavras-chave: mineração de texto, visualização em grafo, compreensão textual, letramento

Abstract

To know how to read and write a small sentence about one's own life was once considered enough to say that a person is literate. Nowadays, literacy is a much more complex concept than simply to be able to write or read sparse sentences. It is also important to understand what is written and be able to critically analyze it. It is not uncommon to find students with reading comprehension problems. Such problems interfere with these students' ability to learn different subjects and to deal with more complex ideas. To help students to improve reading comprehension, we propose in this paper the use of a text mining tool called Sobek. The tool is capable of extracting the main concepts from a text as well as providing a concise view of the relationship between those concepts. Experiments show that a list of concepts extracted by Sobek from texts is highly accepted as accurate by experts in the subject field.

Keywords: text mining, graph representation, reading comprehension, literacy

1. Introdução

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

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 uma notícia de um 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 experiência própria, a habilidade de usar os textos 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 tomar decisões (Jenner 2003).

As dificuldades apresentadas pelos alunos na compreensão textual e produção escrita são problemas bastante sérios que vem sendo discutido 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 do *PISA (Programme for International Student Assessment)*, colocando o país nas últimas posições dentre aqueles que participam da avaliação (OECD 2013). Tais dados mostram a necessidade de desenvolvimento de métodos e ferramentas que possam apoiar os estudantes em atividades envolvendo a produção textual, presentes em grande parte das disciplinas oferecidas em nível superior.

Aprender a ler e compreender textos é uma etapa fundamental do desenvolvimento da criança. A leitura é uma habilidade essencial, requerida para a maior parte das disciplinas escolares e para a comunicação social. Dificuldade em leitura, escrita e matemática são diretamente relacionadas com baixo desempenho acadêmico (Korhonen et al. 2014). Além disso, dificuldade em leitura e compreensão textual estão associados a dificuldades apresentadas por estudantes em diversas disciplinas (Dirks et al. 2008).

Estudantes com dificuldade de leitura costumam apresentar problemas em relação à compreensão e organização de informações contidas em textos. Esses problemas têm sido abordados através do desenvolvimento de ferramentas voltadas a ressaltar palavras chave de textos (Davis and Braun 2010) e organizadores gráficos que permitem uma nova visualização das informações contidas neles (Rello et al. 2012). Nandhini and Balasundaram (2013) propuseram uma aplicação que extrai conceitos chaves do texto e aplicaram essa proposta à textos educacionais, com resultados positivos para o apoio à leitura e à retenção de informações pelo leitor. Com relação ao uso das tecnologias de informação e comunicação (TIC) para apoio aos processos de leitura e escrita, diferentes ferramentas têm sido desenvolvidas e empregadas. (Wei et al. 2012) desenvolveram uma ferramenta que integra diferentes estratégias de leitura em um sistema de e-book, demonstrando como a compreensão de textos pode ser melhorada por meio do emprego da ferramenta. (Villalon et al. 2008) implementaram um editor de texto que busca instigar a reflexão por meio do questionamento sobre o conteúdo do texto, sua estrutura e coerência.

Tais pesquisas e seus resultados positivos serviram de influência para a proposta de utilização de mineração de texto com o propósito de auxiliar a compreensão textual. Este artigo propõe a utilização de um minerador de textos chamado *Sobek* para apoio ao letramento e à compreensão textual. O *Sobek* é um minerador de textos desenvolvido para fins pedagógicos, cuja uma de suas principais características é a facilidade de uso, não sendo necessário nenhum treinamento prévio e sendo bastante flexível para diferentes

aplicações. Apesar de simples de utilizar, possui uma ampla gama de configurações, possibilitando a personalização da ferramenta para diferentes fins e ambientes. Através da extração de conceitos relevantes ao texto e das relações entre conceitos, espera-se que os estudantes tenham uma melhor compreensão textual daquilo que foi lido. Como primeira etapa desse estudo, este artigo contém resultados preliminares que comprovam a eficiência do minerador Sobek para a atividade de extração de conceitos e uma proposta para sua utilização no apoio à compreensão textual.

2. Compreensão Textual

Compreensão de textos pode ser entendido como um processo hierárquico de ações e habilidades realizadas pelo cérebro (Tankersley 2003). Essas ações identificam letras e sons, dando significado ao que está sendo lido. A leitura é apenas parte desse processo, que envolve ações como decodificação de símbolos e fonemas. Compreensão textual requer, além de boas práticas de leitura, um entendimento daquilo que se está lendo, dando um significado à coleção de letras e símbolos, e permitindo ao leitor associar um significado às palavras e fonemas. 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 estimular apenas um deles não é o suficiente para garantir que o leitor desenvolva a habilidade com sucesso. Para (Pressley 2000), compreensão textual é a parte principal do letramento.

Proficiência em leitura e escrita é fundamental para o sucesso, seja ele pessoal ou profissional. Estudantes que tem dificuldade em leitura tem risco de obter baixo rendimento na área escolar e ter problemas sociais e de comportamento. Aproximadamente 75% dos estudantes com alguma dificuldade em aprendizagem manifestam problemas sociais que não estão presentes em estudantes que não possuem dificuldades (Sridhar and Vaughn 2001). Muitas vezes, a dificuldade em aprender está associada a dificuldades em letramento. De acordo com Weinstein (2003), estudantes com problemas de aprendizado precisam de constante apoio e utilizam muito mais energia para realizar tarefas consideradas ordinárias por alunos sem problemas de aprendizado. Isso afeta a interação deles com as demais pessoas e torna o aprendizado um processo muito mais desgastante e difícil (Weinstein and Siever 2003).

De acordo com (LaBerge and Samuels 1974), saber ler uma palavra corretamente não garante que a pessoa entende o que está lendo. Isso decorre do fato da leitura de palavras desconhecidas ou pouco utilizadas requererem um esforço cognitivo muito grande, deixando a parte de reconhecimento e compreensão da palavra em segundo plano. Na educação infantil, é possível que as crianças atentem em demasia à correta pronúncia da palavra e não deem atenção ao seu significado junto ao texto. Isso levou os autores a concluir que é importante não apenas ensinar a ler como também ensinar vocabulário e reconhecimento de palavras. Leitores habituais conseguem decodificar as palavras rapidamente, deixando a maior parte de sua atividade cognitiva para a compreensão do que está sendo lido.

(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, ensinando o primeiro grupo a reconhecer diversas palavras escritas e o segundo grupo o significado semântico e associações dessas mesmas palavras.

Após essa etapa, os estudantes leram um pequeno texto que continha as palavras ensinadas e responderam questões sobre o texto. Os estudantes que aprenderam a reconhecer as palavras escritas responderam mais questões corretamente que aqueles que aprenderam apenas o significado das palavras. Esse estudo foi confirmado posteriormente por (Breznitz 1997), demonstrando assim a importância do reconhecimento de vocabulário escrito na compreensão textual.

Alguns estudos tentam encontrar características que diferenciam bons leitores (pessoas com alto grau de compreensão textual e facilidade de leitura) e maus leitores. Entre essas características, pode-se destacar: capacidade de reconhecer as informações que se deseja extrair do texto; habilidade de perceber a estrutura do texto; associação de ideias do texto com informações conhecidas; capacidade de prever o desenrolar do texto; revisão de informações conflitantes (sejam informações contidas no texto ou aquelas obtidas previamente pelo leitor); e o reconhecimento de vocabulários e acentuação (Pressley and Afflerbach 1995). Estudantes jovens ou com pouca prática de leitura tendem a não realizar nenhum dos processos listados previamente. Entretanto, é possível estimular esses processos e desenvolver bons leitores através de abordagens que instiguem a reflexão sobre o texto. Entre essas abordagens destacam-se o questionamento acerca do texto e suas ideias principais, a construção de imagens mentais do texto, a sumarização do texto e a análise da relação dos personagens e fatos ocorridos no texto (Moats 1998; Pearson et al. 2000; Block and Pressley 2002). A maioria dos leitores que foram explicitamente ensinados estratégias de compreensão textual são mais propícios a aprender, desenvolver e utilizar essas estratégias de forma espontânea (Block and Pressley 2002).

Uma forma de auxiliar leitores com dificuldade de compreensão textual é reforçar a visualização da ortografia das palavras. Quando um leitor reconhece uma palavra, ele associa a ortografia dela com a sua fonologia. Isso forma conexões sinápticas que são reforçadas à medida que a palavra é lida novamente. (Ehri 1979; Rosenthal and Ehri 2008) demonstraram que a visualização de novas palavras melhora a memória de crianças tanto do ponto de vista de vocabulário oral quanto do significado destas palavras, além de aprimorar a leitura e compreensão dessas palavras em textos.

3. Minerador de textos Sobek

O Sobek é um aplicativo de mineração de texto desenvolvido com o objetivo de auxiliar professores nos problemas enfrentados no ensino a distância (EaD). Diante de uma crescente quantidade de textos escritos em cursos a distância e um aumento expressivo no EaD nos últimos anos, percebeu-se que havia a necessidade de auxiliar professores que ministravam tais cursos a classificarem e analisarem textos de maneira rápida e automática. Ao fornecer aos professores uma ferramenta para a análise de mensagens e contribuições dos estudantes, o Sobek torna-se útil no apoio ao trabalho dos professores, dando a eles mais tempo para se concentrarem em problemas específicos de suas áreas.

O Sobek contém uma interface gráfica para apoio ao usuário. Ele extrai informações de um texto e as apresenta como um grafo, onde os principais conceitos extraídos dos textos estão dispostos nos vértices e a ligação entre os conceitos são as arestas entre os vértices. Isso permite representar o texto de forma gráfica e auxilia o usuário a visualizar os principais conceitos do texto e suas relações. A simplicidade para realizar

a mineração de um texto é uma das características mais importantes do software.

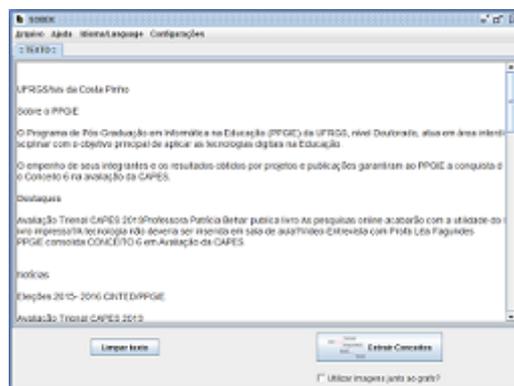


Figura 1. Tela inicial do minerador de textos Sobek

Uma vez que o Sobek tenha completado a extração de conceitos é necessário apresentar os dados extraídos ao usuário. Essa apresentação deve ser clara e objetiva, tornando possível identificar rapidamente os conceitos considerados mais importantes do texto e as demais informações extraídas. O formato de representação gráfica do texto é o formato de grafo. A escolha pelo formato de grafo se deve pela fácil e agradável visualização, possibilidade de conexão entre conceitos e a possibilidade de adicionar informações (como, por exemplo, o número de ocorrências de um conceito) sem a necessidade de alterar a visualização.

O grafo resultante da mineração de texto apresenta os conceitos considerados mais relevantes de acordo com a frequência de ocorrência deles. Também são apresentadas as conexões entre os conceitos. Estas conexões representam relações existentes no texto e podem indicar efeito de causa e consequência, relações temporais ou mesmo conceitos relacionados pelo seu significado. O grafo possui ferramentas que possibilitam sua edição, permitindo ao usuário adicionar novos conceitos, remover conceitos que o usuário não considere adequado ou adicionar e remover ligações entre os conceitos. A edição do grafo é um processo que incentiva a reflexão sobre o texto e permite uma análise mais detalhada dos conceitos e suas relações.

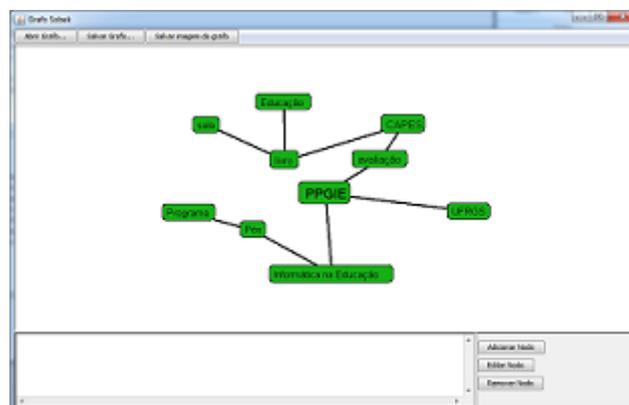


Figura 2. Representação gráfica da operação de extração de conceitos do Sobek.

Outra importante característica presente na representação das informações ex-

traídas do texto é a capacidade de mostrar ao usuário o número de ocorrências de cada conceito e as frases do texto onde este conceito aparece. Esta característica tem por objetivo permitir que o usuário compreenda melhor informações referentes a conceitos específicos, tais como frases relacionadas a esse conceito, seu significado, qual seu papel no texto e quão relevante é frente aos demais conceitos apresentados.

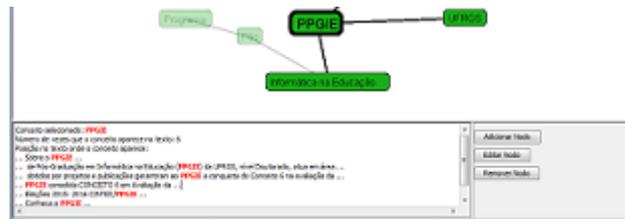


Figura 3. Apresentação das informações do conceito selecionado no grafo.

O algoritmo utilizado no Sobek pode ser dividido em três estágios. O primeiro deles consiste em identificar os conceitos mais relevantes do texto e organizar eles. No caso do minerador de texto Sobek, um conceito será tão importante quanto o número de vezes que ele aparecer no texto. O segundo estágio é a criação de relações entre os conceitos considerados relevantes. O terceiro e último apresenta o grafo de conceitos e relações ao usuário. A seguir, cada um dos estágios será detalhado.

3.1. Identificação dos conceitos mais relevantes do texto

O primeiro passo para a identificação dos conceitos relevantes do texto é separar ele em palavras. O Sobek utiliza espaços em branco, pontuação e sinais de marcação para isso. As palavras são então mapeadas em conceitos, que podem ser *conceitos simples* (conceitos que contém apenas uma palavra) ou *conceitos compostos* (conceitos que contém duas ou mais palavras). Esse mapeamento de palavras em conceitos é realizado através de um processo estatístico, que verifica a frequência com que cada palavra é encontrada 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 por uma série de conceitos simples, sendo assim criado um conceito composto (por exemplo, o conceito composto “meio ambiente” não possui o mesmo significado que a junção dos conceitos simples “meio” e “ambiente”).

Durante o processo de identificação de conceitos, um grupo de palavras chamadas de “*stop words*” é utilizado para remover aquelas palavras que não adicionam informação ao usuário ou não são relevantes ao entendimento do texto. Este conjunto está basicamente formado por artigos e proposições. Após identificar todos os conceitos possíveis de serem extraídos do texto, um processo de “*stemming*” é utilizado para remover redundância e conceitos com mesmo significado conjugados em tempos verbais diferentes ou no plural.

A última parte deste estágio consiste em selecionar os conceitos que farão parte do grafo final. O número de conceitos relacionados pode ser decidido pelo usuário, mas de acordo com (Novak and Cañas 2006) não é necessário mais de 25 conceitos para representar o tema central de um texto. Neste intuito, a configuração padrão do Sobek é de apresentar cerca de 20 conceitos ao usuário, representando assim as informações que aparecem com maior frequência no texto.

3.2. Relacionamento entre conceitos

O segundo estágio no processo de mineração de texto é identificar os relacionamentos entre os conceitos. Um relacionamento entre conceitos é formado quando eles estão próximos um ao outro no texto. Isso pode representar diversos tipos de informação sobre estes dois conceitos, tais como relação de causa e consequência, uma relação temporal ou mesmo que os conceitos têm significados relativos um ao outro. A análise do texto relaciona dois conceitos quando estes estão distantes não mais que z conceitos um do outro e quando não há um *ponto final* entre eles.

Cada conceito pode estar relacionado com mais de um o que poderia produzir um grafo representativo do texto com muitas relações e difícil de ser compreendido. Para reduzir o número de relações que um conceito pode ter e apresentar apenas as mais relações mais significativas, um número máximo de r relações é permitido para cada conceito. O número exato de ligações que um conceito c pode ter depende de sua frequência em relação à frequência do conceito mais recorrente no texto. Isso permite que conceitos mais recorrentes (e, teoricamente, mais relevantes ao texto) possuam mais conexões que os demais conceitos e se destaquem em posições centrais do grafo.

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 encontrados com base em avaliações de usuários e análise de dados. Um valor de z maior que 5 costuma apresentar relações de pouca importância entre conceitos; um valor de r maior que 7 produz grafos com um número demasiadamente alto de relações, podendo até mesmo produzir grafos onde todos os conceitos se relacionam entre si.

3.3. Visualização do grafo de conceitos

O estágio final do Sobek é a construção do grafo utilizando as informações extraídas do texto. Neste grafo, os conceitos 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, baseados na frequência do conceito que ele representa. Quanto maior e mais escuro for o nodo (variando em tons de verde), maior a frequência de ocorrência desse conceito 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.

4. Validação da ferramenta de mineração Sobek

O minerador de textos Sobek extrai conceitos considerados por ele relevantes ao texto. A fim de validar essa extração de conceitos e demonstrar que ela corresponde ao que seria uma sumarização realizada por especialistas, foi realizado um conjunto de testes com o objetivo de medir a “qualidade” dos dados extraídos pelo Sobek. Para esta validação foi requisitado o auxílio de especialistas na área de educação e foram utilizados textos relativos à área de educação. 2 textos foram utilizados nos experimentos, cada um dele analisado por 3 especialistas. Todos os textos, ferramentas e dados desta validação estão disponíveis no site do Sobek ¹ e os resultados podem ser reproduzidos livremente.

¹<http://sobek.ufgrs.br/renote15>

Para avaliar a qualidade das informações fornecidas pelo Sobek, utilizamos 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.

Ao validarmos uma ferramenta de extração de conceitos investigamos o quão próximo de um *gold standard* ela está. Foram utilizados textos referentes a áreas educacionais e, portanto, especialistas dessas áreas devem ser capazes de identificar nesses textos os conceitos mais relevantes e seu enfoque principal. Os conceitos apontados como relevantes pelos especialistas são aqueles que forma a lista *gold standard*. Essa lista servirá de métrica para a avaliação dos conceitos extraídos pelo Sobek.

As listas de conceitos considerada a *gold standard* e aquela formada pela mineração de dados do Sobek para o primeiro artigo apresentado aos especialistas são compostas pelos seguintes conceitos:

$Lista_Gold_Standard_1 = \{benefits, bus, internet, learning, mobile\ devices, network, school, students, technology\ in\ education, wireless, internet\ access, mobile, opportunities, technology, wi-fi\}$

$Lista_Sobek_1 = \{bus, internet, learning, mobile, network, school, students, technology, wireless, wi-fi\}$

Para a análise do teste *gold standard* é possível considerar a *sensibilidade* da lista fornecida pelo Sobek e sua *precisão*. A *sensibilidade* é calculada pelo número de termos considerados relevantes que o Sobek corretamente identificou. O valor da *sensibilidade* para este primeiro artigo é $\frac{10}{16} = 62,5\%$. A *precisão* é o número de termos corretamente identificados pelo Sobek em relação ao número total de termos que há na sua lista. Neste cenário, a *precisão* da lista fornecida pelo Sobek é de 100%! Todos os conceitos considerados relevantes pelo Sobek também são considerados relevantes pelo teste de *gold standard*.

As listas de conceitos criada pelo resultado de *gold standard* e fornecida pelo Sobek para o segundo artigo são:

$Lista_Gold_Standard_2 = \{account, children, classroom, communication, communication\ tool, facebook, learning, Parent, school, social\ media, student, teacher\}$

$Lista_Sobek_2 = \{classroom, communication, Facebook, learning, parent, school, social\ media, student, benefits, Sherry\}$

É possível computar os valores de *precisão* e *sensibilidade* da mesma forma que foi realizado para o primeiro artigo. Neste segundo cenário, a *sensibilidade* encontrada foi de 69% e a *precisão* de 90%.

5. Conclusão

A compreensão textual é um importante aspecto da educação do aluno. Se não for propriamente aprimorada na infância pode trazer problemas futuros para a pessoa tanto na área pessoal quanto profissional. Diversos estudos demonstram que é possível auxiliar no processo de compreensão textual através da sumarização de texto e visualização de conceitos chaves. Estes mecanismos auxiliam o leitor a compreender mais facilmente o texto e realizar uma análise crítica daquilo que foi lido.

O editor de texto Sobek, apresentado neste trabalho, poderá ser empregado de forma a ampliar a reflexão feita pelo aluno a partir de sua leitura. Estes momentos de reflexão e de análise do texto, focalizando na compreensão dos principais conceitos abordados e de seus relacionamentos, permitem ao aluno investigar o texto de maneira mais profunda, construindo pouco a pouco a rede de relações necessárias para poder estruturar sua própria compreensão. Através da apresentação de um grafo de conceitos e relacionamentos, é possível que os estudantes compreendam melhor o texto e encontrem associações que passaram despercebidas durante a leitura.

Este artigo propõe a utilização do minerador de textos Sobek para o apoio à compreensão textual. Apoiado em estudos e pesquisas que demonstram o impacto de diferentes técnicas de leitura e do uso de tecnologias no letramento de crianças e jovens, é proposto que a extração de conceitos e visualização oferecidos pelo Sobek possa servir de auxílio aos estudantes. Também é demonstrado que o foco da ferramenta (extração de conceitos chaves de um texto) é bem sucedido e que os conceitos extraídos refletem aqueles considerados importantes por especialistas. O prosseguimento deste trabalho se dará na forma de utilização da ferramenta em experimentos com estudantes a fim de averiguar se há ganho de compreensão textual por parte dos mesmos ao utilizar o Sobek e como este ganho influencia no letramento dos estudantes.

Referências

- Beers, K. (2003). *When kids can't read. What teachers can do*. Portsmouth, NH: Heinemann.
- Block, C. and Pressley, M. (2002). *Comprehension Instruction: Research-based Best Practices*. Solving problems in the teaching of literacy. Guilford Press.
- Breznitz, Z. (1997). Effects of accelerated reading rate on memory for text among dyslexic readers. *Journal of Educational Psychology*.
- Davis, R. D. and Braun, E. M. (2010). *The gift of dyslexia: why some of the smartest people can't read ... and how they can learn*. Souvenir Press.
- Dirks, E., Spyer, G., van Lieshout, E. C., and de Sonnevile, L. (2008). Prevalence of combined reading and arithmetic disabilities. *Journal of Learning Disabilities*, pages 460—473.
- Ehri, Linnea C.; Wilce, L. S. (1979). The mnemonic value of orthography among beginning readers. *Journal of Educational Psychology*.
- Jenner, J. (2003). A bridge to reading and writing literacy: Developing oral language skills in young children.
- 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.
- LaBerge, D. and Samuels, S. J. (1974). Towards a theory of automatic information processing in reading. *Cognitive Psychology*, 6:292–323.
- Marcuschi, L. A.; Xavier, A. C. (2005). Hipertexto e generos digitais: novas formas de construção do sentido.

- Moats, L. (1998). *Reading, spelling, and writing disabilities in the middle grades*. Learning about learning disabilities (2nd ed.), San Diego, CA: Academic Press.
- Nandhini, K. and Balasundaram, S. (2013). Improving readability through extractive summarization for learners with reading difficulties. *Egyptian Informatics Journal*, 14(3):195 – 204.
- 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.
- OECD (2013). Pisa 2012 results: What students know and can do. student performance in mathematics, reading and science (volume i). 1.
- Pearson, P., Kamil, M., Barr, R., and Mosenthal, P. (2000). *Handbook of Reading Research*. Number v. 3 in Handbook of Reading Research. Lawrence Erlbaum Associates.
- Pressley, M. (2000). *What should comprehension instruction be the instruction of?* Handbook of reading research, Volume 3.
- Pressley, M. and Afflerbach, P. (1995). *Verbal protocols of reading: the nature of constructively responsive reading*. Lawrence Erlbaum Associates.
- Rello, L., Saggion, H., Baeza-Yates, R., and Graells, E. (2012). Graphical schemes may improve readability but not understandability for people with dyslexia. *NAACL-HLT*.
- Rosenthal, J. and Ehri, L. C. (2008). The mnemonic value of orthography for vocabulary learning. *Journal of Educational Psychology*.
- 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.
- Tan, A. and Nicholson, T. (1997). Flashcards revisited: Training poor readers to read words faster improves their comprehension of text. *Journal of Educational Psychology*.
- Tankersley, K. (2003). *The threads of reading: Strategies for literacy development*.
- Vieira, M. C. T. (1981). Levantamento das dificuldades de alunos do 1º 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.
- 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.