
O Emprego da Ferramenta de Mineração de Textos SOBEK como Apoio à Produção Textual

Miriam Klemann¹, Eliseo Reategui^{1,2}, Alexandre Lorenzatti³

¹ PPGEDU, UFRGS, Av. Paulo Gama, 110 - Porto Alegre/RS - 90040-060 - Brazil

² PPGIE, UFRGS, Av. Paulo Gama, 110 - Porto Alegre/RS - 90040-060 - Brazil,

³ PPGC, UFRGS, Av. Bento Gonçalves, 9500 - Porto Alegre/RS - 91501-970, Brazil

{miriamklemann, eliseoreategui, alorenza}@gmail.com

Abstract. *This paper describes a proposal for the use of a text mining tool to support text writing. The tool enables the extraction of concepts and relationships from the texts analysed, helping students to evaluate these concepts and to understand better the texts. A case study is presented in which the tool has been used by a student in order to write a text in the area of business accounting. The study showed the potential of the mining tool to support text production, as well as it showed problems that have to be solved in order to enable a more efficient use of the tool by students.*

KEYWORDS: *text writing, text mining*

Resumo. *Este artigo descreve a proposta de utilização de uma ferramenta de mineração de textos como apoio à produção textual. A ferramenta permite a extração de conceitos e de relacionamentos a partir de textos analisados, auxiliando os alunos na avaliação destes conceitos e compreensão mais aprofundada dos textos. Um estudo de caso é apresentado no qual a utilização da ferramenta foi acompanhada no contexto da produção de um texto na área de contabilidade empresarial. O estudo mostrou a potencialidade da ferramenta no apoio à produção textual, bem como problemas que devem ser solucionados para que esta possa ser empregada por mais eficácia pelos estudantes.*

PALAVRAS-CHAVE: *construção textual, mineração de textos*

1. Introdução

A leitura e a produção textual são comumente tomadas como atividades essenciais no desenvolvimento intelectual e social dos estudantes. Entretanto, é comum a dificuldade destes em se expressar, tanto de forma oral quanto escrita. Por exemplo, Luft (1999, p.43) aponta que “(...) os jovens não sabem falar, ... não sabem escrever, a nova geração não tem vocabulário (...)”.

Considerando os problemas que envolvem a expressão principalmente escrita do aluno, dadas as dificuldades relacionadas à própria língua e à necessidade de orientação para a construção de textos, este projeto propõe a utilização de uma ferramenta computacional para apoio à produção textual.

Conforme Bronckart (2003, p.72), “os textos são produtos da atividade humana e, como tais estão articulados às necessidades, aos interesses e às condições de funcionamento das formações sociais no seio das quais são produzidos.” A escrita é um processo histórico que nos acompanha desde a antiguidade, sendo assim, muito antes de existirem e serem usados os computadores, já existiam estudos sobre o desenvolvimento cognitivo e o processo de escrita de crianças, jovens e adultos. Esses estudos sobre a linguagem escrita são válidos até hoje porque nos dão a base para pensar sobre a escrita

antes e depois do advento das novas Tecnologias de Informação e Comunicação (TIC). As novas possibilidades de escrita e de busca de informações nesta nova era da comunicação nos fazem repensar as metodologias que contribuem para este desenvolvimento, bem como, nos desafiam na busca de ambientes de aprendizagem interativos, dinâmicos e significativos.

Este artigo apresenta uma nova abordagem para apoio à construção textual baseada na utilização de uma ferramenta computacional capaz de extrair grafos a partir da captura dos principais termos utilizados no textos e seus relacionamentos. Uma metodologia para utilização da ferramenta é proposta e apresentada juntamente com um estudo de caso. A última seção do artigo apresenta considerações finais e propostas de trabalhos futuros.

2. Produção textual

Para Marcuschi (2008, p. 77) a produção textual se assemelha com um jogo, pois antes de iniciar temos que conhecer algumas regras, o espaço onde vai acontecer, qual o papel e função dos atores envolvidos. Também envolve algumas decisões e são utilizadas variadas estratégias.

As atividades de produção textual sempre têm um objetivo resultando em diferentes textos, com enfoques e estruturas diferentes. Escrever uma história, por exemplo, é diferente de produzir um artigo para jornal, um poema ou rimas, ou seja, é preciso escrever o texto de acordo com o gênero e tipo textual. Além disso, é necessário preocupar-se também com o sucesso dos objetivos da produção textual, como a interação entre o produtor do texto e o seu receptor. Escrever é diferente de falar. A linguagem escrita existe, conforme Rizzo (1986), porque existe linguagem oral, que é a necessidade social de comunicação entre os indivíduos. Portanto a escrita reflete a linguagem oral podendo ser expressa de forma clara, para si mesmo e para os outros, pois todo texto tem um autor que fala e escreve (sujeito) e projeta ao executar o texto, a sua intenção (Demo, 1996).

Bakhtin (2000) afirma que o texto é uma construção de signos criada por um indivíduo para expressar suas opiniões, valores e crenças, resultantes da sua interação com outros indivíduos, inseridos em uma cultura e dentro de um período histórico. Se esse ser social se expressa através do texto que produz, deseja também que alguém o compreenda. O texto não apenas revela significados construídos pelo processo dialógico entre sujeitos, mas que, para compreendê-lo, só será possível no encontro dialógico entre sujeitos.

Diversas atividades podem intervir no texto do aluno para progressivamente aproximá-lo da escrita mais elaborada. Essas atividades podem ser desenvolvidas desde o Ensino Fundamental, passando pelo Ensino Médio até o Ensino Superior, através de produções envolvendo várias atividades diferenciadas. Um dos desafios educacionais é acompanhar a dinâmica dos tempos modernos, por meio da integração com os novos recursos de linguagem, a exemplo dos editores de textos para produções textuais mais elaboradas e sobre a influência que o texto exerce nos relacionamentos sociais. Koch (1998a) sustenta:

“(...)É preciso pensar a linguagem humana como lugar de interação, de constituição de identidades, de representação de papéis, de negociação de sentidos. É preciso encarar a linguagem humana não apenas como representação do mundo e do pensamento ou como instrumento de comunicação, mas sim, acima de tudo, como forma de interação social. (...)”

Cabral (2009) enfatiza que freqüentemente os alunos não apresentam dificuldades em se expressar na oralidade através da linguagem coloquial. No entanto,

os problemas aparecem quando surge necessidade de produção textual. A autora apresenta quatro elementos para escrever um texto: [1] a repetição (retomadas de elementos feita por pronomes ou por palavras e expressões equivalentes ou sinônimas), [2] a progressão (acrescentar novas informações ao que já foi dito complementando a repetição buscando equilibrar o que já foi dito com o que se vai dizer, garantindo a continuidade do tema), [3] a não-contradição (não devem surgir elementos que contradigam aquilo que já foi citado) e [4] a relação (os fatos e conceitos devem estar relacionados). Todas essas partes compõem o texto, elas surgem uma após a outra, relacionando-se com o que já foi dito ou com o que se vai dizer e podem ajudar a avaliar o grau de coesão dos textos. Val (1999, p.7) define a coerência como o nexos, a ligação entre os conceitos. A coesão é definida pelo autor com relação à expressão desse nexos no plano lingüístico. O professor, por sua vez, pode localizar e identificar problemas no texto e propor correções, atividades e novas situações para alcançar os objetivos propostos.

Neste artigo, propomos a utilização de uma ferramenta de mineração de texto como forma de explorar estes elementos que fazem parte da construção textual. Na metodologia proposta, os próprios estudantes podem desenvolver formas de construção e produção de escrita de textos a partir dos resultados obtidos com a ferramenta computacional. Neste contexto, os estudantes podem também avaliar sua produção textual – coesão e coerência e desenvolvimento a partir das formas de representação visual apresentadas pelo programa, detalhado na seção 4 deste artigo.

Com relação à compreensão dos textos, Demo (1996) argumenta:

“(...) compreender o sentido de um texto implica estabelecer relações entre texto e significado, colocar em movimento modos de entender e compreender, indagar possibilidades alternativas de compreensão, perceber e dar sentidos. Esta dinâmica avança ainda mais, quando se trata de saber fazer e refazer um texto, passando-se de leitor a autor. Aparecendo a elaboração própria, torna-se visível o saber pensar e o aprender a aprender (...)”

Estas situações de aprendizagem são atividades diversas, planejadas e organizadas a partir da observação do processo de aprendizagem dos alunos, em função de algum material/texto específico, a partir de propostas dos alunos, de alguma situação ocorrida na interação estabelecida na aula, ou até mesmo (re)planejadas a partir de nossa reflexão sobre a prática. Para Piaget (1972), nenhum conhecimento é transmitido, mas sim, reconstruído internamente pelo sujeito a partir da reorganização de elementos tirados de estruturas anteriores. Para este autor (1995), podem-se distinguir três tempos neste processo: abstração (que consiste em compreender certos modos de organização dos conhecimentos do sujeito), reflexionamento (que torna a projetar o que foi abstraído em um plano de conhecimento superior) e a reflexão (ou reconstrução em um novo plano). Numa produção textual podemos observar o resultado deste processo, pois o aluno inicialmente considera isoladamente as informações para depois registrar, em linguagem natural, as relações que julga importantes e/ou relevantes às informações que obteve.

Johnson (2001) relata que nos anos que ainda escrevia com caneta e papel, ou usando uma máquina de escrever, elaborava cada frase antes de começar a escrevê-la. Tudo era planejado antes, pois alterar a seqüência das palavras e/ou frases era possível, mas tornava o texto sem muito sentido, difícil de ler. Era preciso ser capaz de reter toda a seqüência de palavras, o que significava que a mente tendia naturalmente para uma sintaxe mais simples, mais direta. Neste sentido, o meio digital trouxe facilidades e novas possibilidades na elaboração dos textos. Podemos observar uma mudança na forma da organização e da seqüência do pensamento que se reflete nitidamente nas frases elaboradas detalhadamente e cheias de significados.

3. Construção textual apoiada nas novas tecnologias

O uso das novas tecnologias e o domínio dos recursos oferecidos (como as ferramentas de corte, de realce, salvar um arquivo, apagar e reescrever o texto, podendo fazer isso quantas vezes for necessário) podem levar o aluno a compreender como melhorar seu texto. A tecnologia pode interagir de forma sistemática e significativa nas práticas de escrita e leitura no dia-a-dia da sala de aula, fomentando um trabalho de qualidade. A maneira de ler, interpretar e construir textos está sendo alterada e ocupa espaço importante na aprendizagem, porque tem papel decisivo e fundamental na capacidade de reflexão, expressão e criação de novas informações e conhecimentos.

Utilizando programas adequados, o computador pode facilitar o desenvolvimento da expressão, permitindo a realização de trabalhos diversos com imagens, desenhos e textos separadamente ou de forma combinada. As atividades tornam-se muito mais interessantes e um fator que contribui para manter o interesse do aluno pela atividade, é ter oportunidade de refazer, re-escrever, inserir, deletar, copiar, colar quantas vezes achar necessário e ter sempre o retorno imediato.

Johnson (2001) coloca que com o uso do computador, de um processador de textos, muda a maneira de escrever, de conceber as frases, sendo que o processo de pensamento passa a acontecer no desenrolar do processo de escrever. Lembra que no início registrava tudo a mão e somente depois de efetuar uma composição completa de suas idéias, fazia a transcrição para o computador.

Com o passar do tempo, esta rotina também mudou, ocorreu uma "virada", de modo que a caneta e o papel já lhe pareciam uma tecnologia inferior. Então, processos de pensamento e digitação começaram a coincidir. Conforme o autor, sempre é possível acrescentar mais uma parte tornando mais fácil o ato de escrever e mudando a própria substância do que estava escrevendo, trazendo também um impacto ao seu pensamento. O mais interessante para ele é a magia do processador de texto e a mudança na maneira de criar.

Com a entrada dos computadores no mercado, sua popularização e facilidade de uso, aconteceu uma verdadeira revolução na escrita, sobretudo em função das ferramentas de edição de texto, adicionando à escrita maior flexibilidade e eficiência no processo de construção textual. Isso nos remete à possibilidade de, a partir de um texto elaborado, criar muitos outros. Não partindo exatamente de uma idéia apenas.

Para Giombini (2004) existe a necessidade do uso das palavras como uma ferramenta de comunicação do pensamento. A autora cita que as funções que podem ser identificadas na formação de textos escritos são o resultado intencional da combinação de som, sinal e significado. O mesmo ocorre na formulação de um diagrama relacionando estes elementos, pois os símbolos desempenham uma expressão, um sinal e representam algo (significado) criando uma comunicação muito eficaz. Assim como um diagrama, os Mapas Conceituais mostram graficamente as relações entre conceitos e permitem que se visualize, a partir deles, as construções que o sujeito faz mentalmente. Esse exercício de leitura, reflexão e construção de mapas, pode também servir de apoio à produção textual. Conforme a autora, isso requer observar algumas regras de composição, nas quais os conceitos são expressos de forma resumida (ex. palavras, termos) ligados entre si por linhas que narram essas relações.

Apesar do interesse no emprego de mapas conceituais para trabalhar a compreensão de textos, reflexão e produção textual, é instigante a idéia de um processo de construção de textos mais interativo, também fundamentado na análise de conceitos e relacionamentos. A próxima seção apresenta a ferramenta SOBEK como uma proposta de recurso computacional interativo para apoio à produção textual.

4. Ferramenta de mineração de textos – SOBEK

SOBEK (Macedo et al., 2009) é uma ferramenta capaz de extrair termos frequentes em documentos e encontrar os relacionamentos entre estes a partir de um processo conhecido como mineração de texto. Este processo é definido como um método de extração de informações relevantes em bases de dados não estruturadas, ou semi-estruturadas (Feldman e Sanger, 2006). No caso da ferramenta SOBEK, um algoritmo específico definido por Shanker (2003) foi implementado, baseado em análises estatísticas dos textos e representação das informações extraídas em um modelo de grafo chamado de distância n -simples (*n-simple distance*).

A ferramenta foi idealizada como recurso de apoio aos professores no acompanhamento de trabalhos de escrita colaborativa (Macedo et al., 2009). Aqui, a ferramenta foi adaptada para ser utilizada com os próprios alunos – como forma de auxiliá-los no processo de produção textual. A ferramenta opera da seguinte maneira:

- (1) Copia-se o texto que se quer trabalhar para a área de entrada de dados da ferramenta. Alternativamente, carrega-se um texto em formato *txt*, *doc* ou *pdf*. A figura 1 mostra a ferramenta com um texto sobre o tema “a Escrita”¹.

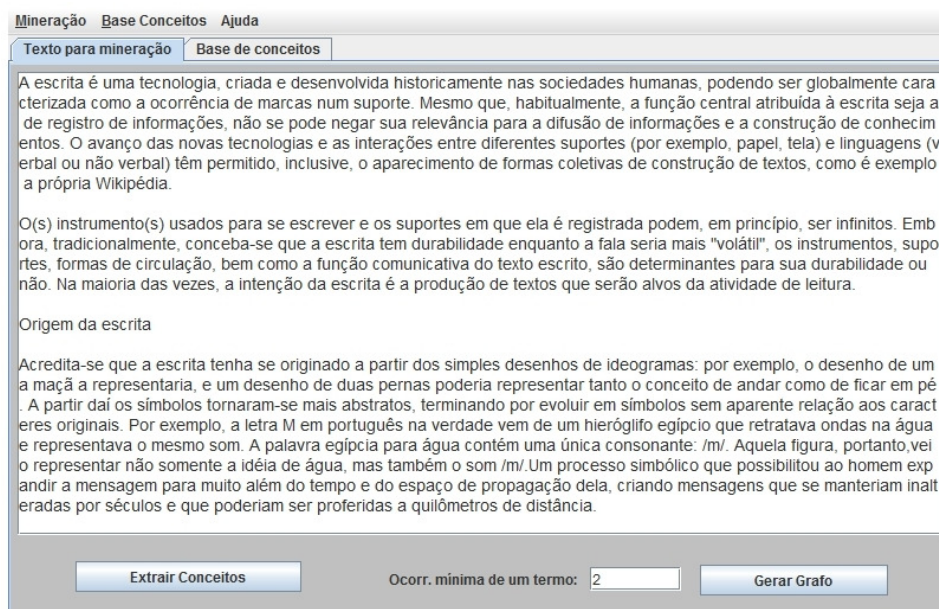


Figura 1 – Ferramenta SOBEK

- (2) A partir deste texto, cria-se uma base de conceitos automaticamente através do processo de mineração de textos (botão “Extrair Conceitos”, figura 1). Estes conceitos correspondem a um conjunto de palavras e termos que ocorrem com mais frequência no texto, excluindo-se artigos, preposições, e outras palavras consideradas sem significado relevante². O parâmetro “ocorrência mínima de um termo” (figura 1) permite que apenas termos com frequência maior que o limite estabelecido sejam considerados na extração de conceitos.
- (3) Extraídos os conceitos, gera-se um grafo com os conceitos mais frequentes e seus relacionamentos (figura 2).

¹ Texto introdutório do artigo “Escrita”, extraído da enciclopédia Wikipedia, disponível no endereço: <http://pt.wikipedia.org/wiki/Escrita>, acessado em agosto de 2009.

² O sistema permite utilizar uma lista chamada de *stop-words* com termos a serem desconsiderados do processo de mineração. Atualmente a ferramenta trabalha com listas de *stop-words* em português e em inglês, mas facilmente pode ser adaptada para outras línguas.

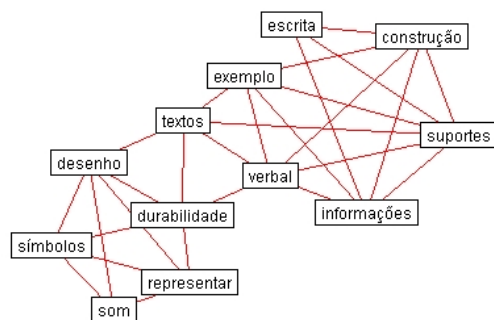


Figura 2 – Grafo gerado para o texto sobre o tema “a Escrita”

A partir do exemplo exposto, percebe-se que os termos extraídos do texto são efetivamente relevantes com relação ao tema tratado. A próxima seção apresenta uma metodologia de utilização da ferramenta SOBEK como apoio à produção textual.

5. A Mineração de Textos como apoio à Produção Textual

A premissa por trás do processo de mineração de textos baseado em métodos estatísticos é de que termos importantes aparecem repetidas vezes no texto. Apesar de muitas vezes correta, a premissa frequentemente falha na medida em que outros termos não tão relevantes podem também ser recorrentes. A partir desta limitação, definimos um processo para produção textual empregando a ferramenta SOBEK, como mostra o diagrama da figura 3.

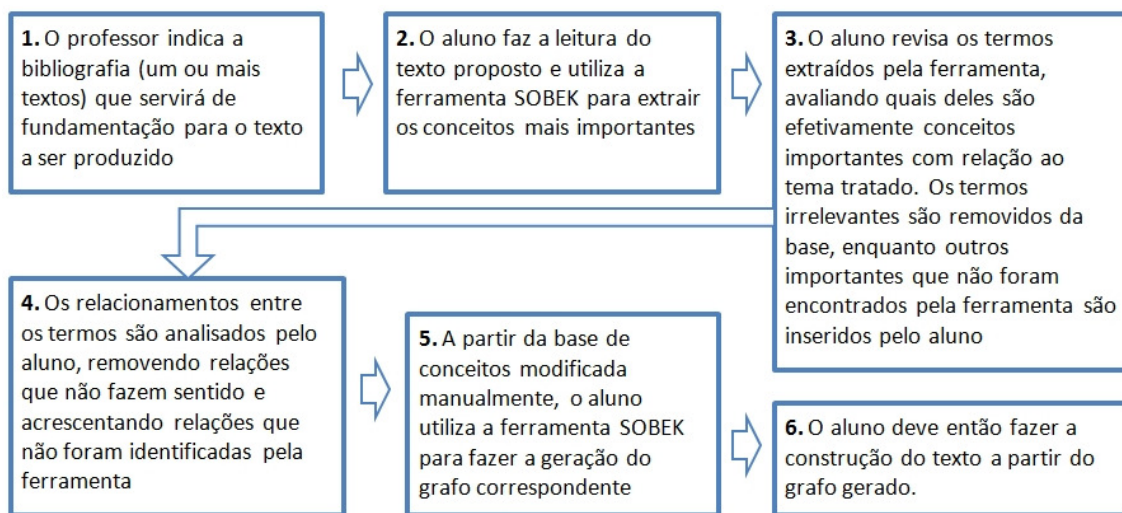


Figura 3 – Processo Interativo para Produção Textual

O processo para construção do texto é caracterizado pela interação que se estabelece entre o aluno e o sistema, inicialmente na construção e refinamento da base de conceitos, e em um segundo momento na criação e ajuste do grafo. Estes momentos de reflexão, de análise do texto focalizando na compreensão dos principais conceitos abordados e de seus relacionamentos, permitem ao aluno penetrar no texto de maneira mais profunda, construindo pouco a pouco a rede de relações necessárias para poder estruturar sua própria narrativa.

5.1. Estudo de caso

Para avaliar o processo de construção textual apoiado pela ferramenta SOBEK, acompanhamos a atividade de construção de texto realizada por um estudante de 20 anos do segundo semestre de Engenharia Mecânica na Universidade de Jaraguá do Sul (UNERJ – SC). Em termos metodológicos, classificamos a investigação realizada como

estudo de caso com observação não participante. Foi solicitado ao aluno que fizesse a leitura de um texto na área de contabilidade empresarial e que elaborasse um resumo desta leitura³.

O estudante iniciou a atividade inserindo o texto fornecido na ferramenta SOBEK. Em seguida, selecionou a opção “Extrair Conceitos” para que a ferramenta selecionasse automaticamente os termos mais freqüentes no texto, e que a partir destes se iniciasse a análise dos principais conceitos do texto. A tela abaixo apresenta os termos mais freqüentes identificados (coluna da esquerda). A coluna do meio mostra os termos relacionados ao termo selecionado na coluna da esquerda, no caso a palavra *anormal*. A coluna da direita apresenta outros termos freqüentes mas que não estão relacionados à palavra selecionada (*anormal*). Utilizando os botões com setas “>>>” e “<<<” é possível acrescentar e remover relacionamentos entre os termos.

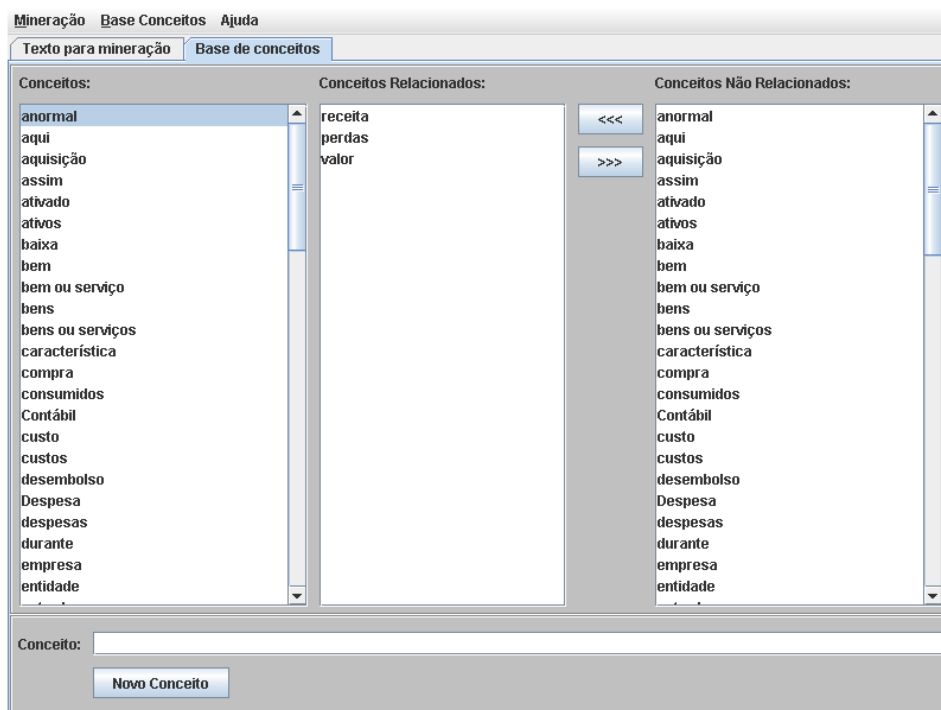


Figura 4 – Termos mais freqüentes no texto

O próximo passo consistiu-se da verificação dos conceitos pertinentes. É possível constatar que na lista de conceitos extraídos pela ferramenta (coluna esquerda, figura 4), muitas das palavras e termos não contêm significado relevante para a compreensão do tema escolhido, i.e. contabilidade empresarial. Com base nesta verificação, o aluno organizou a lista de conceitos, eliminando os termos considerados pouco pertinentes, e acrescentando termos importantes que apareciam no texto e que não foram identificados pela ferramenta. Palavras escritas no singular e plural com os mesmos conceitos relacionados foram excluídas, bem como termos de ligação entre palavras. Também foram observadas algumas repetições nos relacionamentos e incluídos conceitos não extraídos. Neste processo, o aluno voltou ao texto original algumas vezes para releitura, na medida em que surgiram dúvidas sobre os conceitos e suas relações. O aluno se mostrou muito atento na realização da atividade, e refez duas vezes este procedimento para ter certeza de estar tudo correto. É importante observar que este “ir e vir” fez com que o aluno aprofundasse sua compreensão sobre o tema

³ O texto sugerido faz parte da apostila sobre Terminologia Contábil (página 5 até 8) do Centro Universitário de Jaraguá do Sul (UNERJ), curso de Engenharia Mecânica, Prof. Paulo Onildo de Matos.

proposto e percebesse detalhes importantes em relação aos conceitos. No passo seguinte, o aluno faz a construção do grafo (figura 5), selecionando a opção “Gerar Grafo” da ferramenta SOBEK.

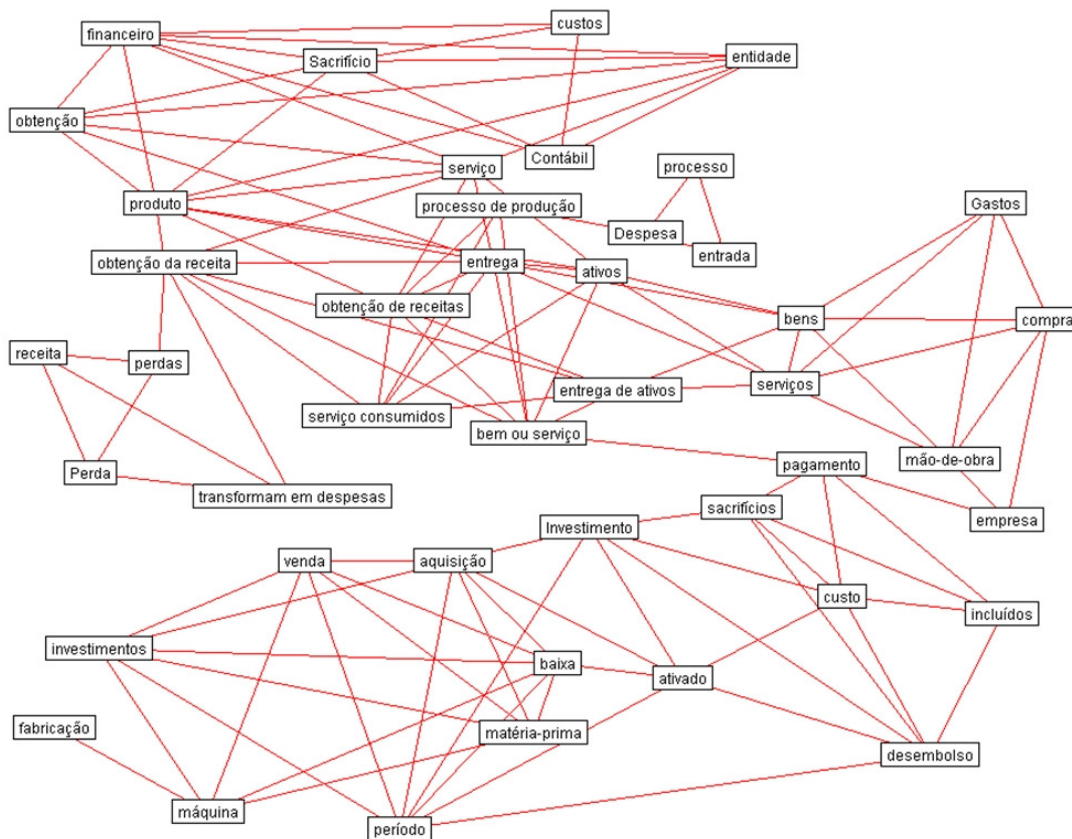


Figura 5 – Grafo gerado pela ferramenta SOBEK

Logo após, ao observar o grafo, o aluno imediatamente identificou os conceitos e relacionamentos apresentados como corretos, passando a descrevê-los oralmente numa seqüência bastante coerente, apresentando seus significados e justificando suas posições no grafo. Por exemplo, ao se referir ao conceito *Gastos*, que aparece ao lado direito do grafo, o estudante percebeu a relação entre os termos *bens* e *serviços* adquiridos, bem como *Gastos com mão-de-obra*, e ainda *Gastos na compra* de um imobilizado, como consta no texto original. No passo seguinte, o aluno procedeu à construção do texto a partir do grafo construído. Como os conceitos relacionados ao tema proposto já haviam sido bastante trabalhados em todo o processo de construção do grafo, a produção de texto feita pelo aluno ocorreu de forma muito natural, espontânea e com definições muito claras.

Abaixo, seguem algumas observações do aluno com relação ao processo de construção do texto apoiada pela ferramenta SOBEK.

“A possibilidade de extrair de um texto as principais informações e conceitos úteis é muito interessante. Mesmo tendo que fazer “uma limpeza” manualmente nos conceitos relacionados a experiência foi muito válida. O fato de ter que realizar e repetir a leitura de algumas partes do texto para ter a certeza que estava correto ajudou a entender e memorizar o conteúdo. Desta forma a interpretação e compreensão dos conceitos foi acontecendo de forma rápida. Após observar que o grafo fazia sentido, a produção do texto (resumo) foi fácil de fazer. Bastava olhar o grafo, acompanhar os relacionamentos e transformar a parte gráfica em texto.”

Percebe-se no relato do aluno a total compreensão do método de produção textual proposto, no qual a construção do texto é realizada através de um processo estruturado e interativo. Neste processo, o aluno interage com a ferramenta Sobek no desenvolvimento de cada uma das etapas de construção textual. O aluno também fez as seguintes considerações com relação a melhorias necessárias na ferramenta:

“O programa deveria sofrer apenas algumas alterações, melhorando a organização das palavras e a definição de linhas de assuntos principais, aqueles ligados diretamente, para um mais rápido entendimento do assunto. E ainda, a opção para que se possa fazer a alteração dos termos diretamente nos grafos.”

A primeira observação do aluno mostra sua inquietação com relação à funcionalidade da ferramenta de extrair conceitos realmente pertinentes. Efetivamente, no procedimento realizado, muitos termos irrelevantes foram identificados pela ferramenta, sobretudo pelo fato do parâmetro “ocorrência mínima de um termo” ter sido mantido em 2 (figura 1) – um valor baixo para um texto de três páginas. Consta-se portanto que essa função não está clara para os alunos, e que um processo mais automático de seleção deste parâmetro deveria ser desenvolvido.

Já seu segundo comentário diz respeito não tanto à funcionalidade disponibilizada de alteração da base de conceitos, mas sim à interface do sistema que poderia ser mais adequada para realização das modificações nos grafos. Esta alteração na interface do sistema, dentre outras, está sendo estudada pela equipe de desenvolvimento da ferramenta SOBEK.

6. Considerações Finais

A principal contribuição deste trabalho foi propor uma forma de utilização de uma ferramenta de mineração de textos para auxiliar os alunos no processo de produção textual. Um estudo de caso foi apresentado, no qual um estudante realizou a leitura de um texto na área de contabilidade empresarial, e depois disso empregou a ferramenta num procedimento que envolveu: extração de termos frequentes; criação de uma base de conceitos e relacionamentos a partir dos termos extraídos automaticamente; geração de um grafo correspondente aos termos e relacionamentos estabelecidos; escrita do texto com base no grafo gerado.

Observou-se que o processo de “idas e vindas” do aluno ao texto, para completar cada uma das etapas de utilização da ferramenta, foi construtivo para promover uma melhor compreensão do texto e do tema tratado. A produção do texto de maneira bastante fluida foi atribuída às várias etapas adotadas no uso da ferramenta.

Com relação à ferramenta, alguns problemas foram observados, como a dificuldade de extração de termos efetivamente relevantes, ou a necessidade de construção de uma interface gráfica para manipulação da base de conceitos diretamente nos grafos gerados.

A idéia de utilização de grafos para apoiar a construção textual é análoga à idéia de emprego de mapas conceituais, abordagem desenvolvida na década de 70 por Joseph Novak (1984). Enquanto ferramenta educacional, os mapas conceituais podem auxiliar os estudantes e os educadores a explicar relações, significados e conceitos dos materiais de aprendizagem. No entanto, como salientado anteriormente, as ferramentas de construção de mapas conceituais exigem que o aluno trabalhe sozinho na construção dos mapas – não permitindo uma construção mais interativa como no caso da ferramenta SOBEK. Neste sentido, seria interessante em um trabalho futuro comparar mais detalhadamente a utilização de ambas as ferramentas como apoio à produção textual. Também estão sendo avaliadas diversas modificações na ferramenta SOBEK para que

esta possa ser empregada com mais facilidade pelos estudantes. Um estudo com um grande número de alunos está sendo estruturado para que se possa avaliar de maneira mais ampla os benefícios introduzidos pela ferramenta no processo de produção textual.

Referências

- BAKHTIN, Mikhail. *Estética da Criação Verbal*. Tradução: Maria Ermantina G. Gomes. São Paulo: Martins Fontes, 2000
- BRONCKART, Jean-Paul. *Atividade de linguagem, textos: por um interacionismo sócio-discursivo*. Tradução: Anna Raquel Machado. São Paulo: EDUC, 2003
- CABRAL, Marina. O Texto Escrito. *Brasil Escola, 2009*. Disponível em: <http://www.brasilecola.com/redacao/texto-escrito.htm>. Acessado em agosto de 2009.
- CASTELLS, Manuel. *A Galáxia da Internet: reflexões sobre a internet, os negócios e a sociedade*. Tradução: Maria Luíza X. de A. Borges. Rio de Janeiro: Jorge Zahar Editor, 2003. 243 p.
- DEMO, Pedro. *Educar pela pesquisa*. Campinas: Autores Associados, 1996.
- FELDMAN, R., SANGER, J. *Text Mining Handbook*. Cambridge, Inglaterra: Cambridge University Press, 2006.
- GIOMBINI, L. Concept Maps: Theory, Methodology, Technology. *Proceedings of the First Int. Conference on Concept Mapping*. Pamplona, Spain 2004.
- JOHNSON, Steven. *Cultura da Interface: como o computador transforma nossa maneira de criar e comunicar*. Trad. Maria Luíza X. de A. Borges; revisão técnica, Paulo Vaz. Rio de Janeiro: Jorge Zahar Ed., 2001.
- KOCH, I. G. V. *A coesão textual*. 10. ed. São Paulo: Contexto, 1998a.
- KOCH, I. G. V. *A coerência textual*. 8. ed. São Paulo: Contexto, 1998b.
- LUFT, Celso Pedro. *Língua e liberdade*. 7.ed. São Paulo: Ática, 1999.
- MACEDO, A., Reategui, E., Lorenzatti, A., Behar, P. Using Text-Mining to Support the Evaluation of Texts Produced Collaboratively. *Education and Technology for a Better World: Selected papers of the 9th World Conference on Computers in Education*, Bento Gonçalves, Brazil. Springer, 2009.
- MARCUSCHI, Luiz Antonio. *Produção Textual, Análise de Gêneros e Compreensão*. Coleção: Educação Lingüística - 1ª Edição. São Paulo: Editora Parábola, 2008.
- NOVAK, J.D., & Gowin, D.B. (1984). *Learning how to learn*. New York, N.Y.: Cambridge University Press.
- PIAGET, J. *Os estágios do desenvolvimento intelectual da criança e do adolescente*. In: Piaget. Rio de Janeiro: Forense, 1972.
- RIZZO, Gilda. *Fundamentos e Metodologia da Alfabetização – Método Natural*. 4ª edição. Rio de Janeiro. Coleção Educação Prática, 1986.
- SCHENKER, A. *Graph-Theoretic Techniques for Web Content Mining*. PhD thesis, University of South Florida (2003).
- VAL, Maria da Graça Costa. *Redação e textualidade*. 2ª Edição. São Paulo: Martins Fontes, 1999.