

Acompanhamento da Interação e Produção Textual Coletiva Através de Mineração de Textos

Monitoring of interaction and collective production of texts through text mining

Resumo:

Este artigo apresenta a ferramenta Rede de Conceitos desenvolvida a partir da tecnologia de mineração de texto. O principal objetivo é extrair e relacionar os termos tratados com maior incidência na produção textual e exibir o resultado através de um grafo. A Rede foi implementada no Editor de Texto Coletivo (ETC) que é uma ferramenta online que permite a produção de textos de forma síncrona ou assíncrona pelos autores. Este artigo descreve a aplicação da Rede tanto em produções textuais coletivas, quanto na ferramenta fórum. O intuito é de oferecer suporte para o professor gerenciar o alto volume de dados oriundos das negociações e construção do texto em um menor tempo que o demandado atualmente. Os resultados apontam que a Rede de Conceitos pode auxiliar o professor, pois possibilita a obtenção de indicadores sobre a qualidade da produção textual. Além disso, ele pode analisar as mensagens postadas nos fóruns, sem necessariamente ler todo o conteúdo deste material anteriormente.

Palavras-chave: Mineração de Texto. Produção textual coletiva.

Abstract:

This paper introduces the Concepts Network tool, which was developed based on text mining technology. The tool's main goal is to extract and relate relevant terms found in texts and present them in the form of a graph. The Network was implemented and integrated to ETC, a collective text editor which enables the production of texts synchronously or asynchronously by different authors. This paper describes the use of the Concepts Network tool in collective text productions as well as in a discussion forum, aiming at providing support for the teacher to manage the large volume of data from students' negotiations and construction of texts in a shorter period of time than would normally required. The results show that the Concepts Network can help teachers obtain indicators on the quality of forums' posts and texts produced, without having to read the entire contents of this material available.

Keywords: Text Mining. Collective textual production.

Alexandra Lorandi Macedo

Breno Azevedo

Patricia Alejandra Behar

Eliseo Reategui

Universidade Federal do Rio Grande do Sul

1 Introdução

O presente artigo tem como foco o Editor de Texto Coletivo – ETC, disponível em <http://www.nuted.ufrgs.br/etc2>, mais precisamente, em uma de suas funcionalidades denominada Rede de Conceitos. O ETC¹ favorece o processo de escrita coletiva a distância onde os sujeitos interagem a partir de um objetivo comum, ou seja, a construção de um texto sobre um determinado tema. Em geral, a produção coletiva costuma se dar a partir de inúmeras interações, trocas, confronto entre diferentes pontos de vistas, construções e reconstruções acerca de um assunto específico. Em paralelo à produção textual propriamente dita, tem-se, por exemplo, ferramentas de

MACEDO, Alexandra Lorandi; AZEVEDO, Breno; BEHAR, Patricia Alejandra; REATEGUI, Eliseo. Acompanhamento da Interação e Produção Textual Coletiva Através de Mineração de Textos. *Informática na Educação: teoria & prática*, Porto Alegre, v. 14, n. 2, p. 33-45, jul./dez. 2011.

1 O ETC proporciona aos usuários recursos semelhantes aos disponibilizados pelo *Google Docs* ou *Wiki*, com a diferença que contempla recursos educacionais que favorecem a mediação, acompanhamento e avaliação do processo por parte do professor.

comunicação, como é o caso do fórum. Esta ferramenta apóia a negociação e discussão sobre a produção textual. Assim, tanto o texto propriamente dito, quanto as negociações acerca deste, geram um alto e permanente volume de dados para serem gerenciados pelo professor que tem a função de acompanhar esse processo de construção. Diante desta perspectiva, o intuito deste estudo é o de oferecer suporte para o professor gerenciar este volume de dados em um menor tempo que o demandado atualmente. Para viabilizar tais condições foi desenvolvida a Rede de Conceitos, uma ferramenta que utiliza técnicas de mineração de textos.

A mineração de textos (MT) é um conjunto de técnicas e metodologias da Ciência da Computação que tem por objetivo extrair informações importantes provenientes de documentos não estruturados. Este tem sido utilizado em diversas áreas do conhecimento e, neste artigo, está focado como forma de auxiliar o professor no gerenciamento dos dados oriundos das trocas coletivas e da produção de texto propriamente dito.

Diante do exposto, apresenta-se na seção dois deste artigo a relação entre a interação e o volume de dados no processo de escrita coletiva. A seção três aborda aspectos técnicos da mineração de textos, as aplicações educacionais e os resultados coletados em duas experiências de uso da Rede de Conceitos. Por fim, a seção quatro traz as considerações finais sobre a aplicação da ferramenta.

2 Escrita Coletiva: a relação entre o processo de interação e o volume de dados

Os recentes e significativos avanços tecnológicos têm-se destacado por permitirem o desenvolvimento de atividades que privilegiam a

troca, a comunicação e a negociação entre os sujeitos participantes. Deste cenário faz parte o processo de escrita coletiva a distância que oferece ferramentas que apóiam a troca de ideias e o confronto de diferentes pontos de vista. Diante desta perspectiva, entende-se que os alunos aprendem em um processo interativo, coletivo, que pode favorecer uma aprendizagem qualitativamente melhor do que a desenvolvida numa prática individual. O autor Ascott (1997, p. 338) destaca que “[...] sem interação nada de novo acontece. Sem interação nenhum significado é gerado. Sem interação nenhuma experiência é criada [...]”. Desta forma, constitui-se o processo coletivo de criação, de construção, de troca. Esse movimento é avesso à inércia e convoca todos os participantes para atuar, para interagir, para se expor e se modificar. Neste contexto, entende-se o coletivo sendo todos os alunos como sujeitos em movimento, inclusive o professor que acompanha e media as produções coletivas.

Em contextos digitais, os conflitos nos grupos costumam ser mais complexos para resolver do que os presenciais. A interpretação que pode ser dada para uma ou outra escrita/argumentação sobre determinado tema, por vezes, cria dimensões difíceis de administrar. Para favorecer a continuidade no processo acredita-se ser importante garantir a liberdade entre os participantes. Tal liberdade compreende momentos para questionar, posicionar-se e contra-argumentar novos temas e novas soluções que poderão emergir dos debates, favorecendo a continuidade das trocas e construção de novos saberes. É nesta perspectiva que está inserido o Editor de Texto Coletivo. Ou seja, o espaço de interação oferecido pelas ferramentas de comunicação e edição de texto, gera um alto e permanente volume de dados e informações, dificultando o acompanhamento destes pelo professor.

Essa contingência requer deste profissional

um precioso tempo que poderia ser dedicado para intervenções pedagógicas junto aos alunos no processo de construção coletiva, com vistas ao avanço e aprimoramento dos mesmos. Assim, a implementação da Rede de Conceitos no ETC visa apoiar a análise e gerenciamento desses dados, pelo professor, num espaço de tempo menor do que o demandado até então. Para tanto, buscou-se recursos na tecnologia de Mineração de Textos, a qual oferece condições de atender as necessidades identificadas neste estudo.

3 Mineração de Textos

Segundo Feldman e Sanger (2007), a mineração de textos pode ser definida como um processo intensivo de conhecimento no qual um usuário interage com uma grande quantidade de documentos utilizando ferramentas para análise dos mesmos. O objetivo é extrair informações úteis a partir de coleções de documentos. As informações úteis são identificadas em padrões interessantes nos dados textuais não estruturados.

Os sistemas de mineração de textos baseiam-se em rotinas de pré-processamento, algoritmos para descoberta de padrões, e elementos para apresentação dos resultados. Nestes sistemas, as operações de pré-processamento baseiam-se na identificação e extração de características representativas dos documentos em linguagem natural. Essas operações são responsáveis por transformar dados não estruturados, armazenados em coleções de documentos, em uma estrutura expressa em um formato intermediário (FELDMAN; SANGER, 2007, HEARST,1999, TAN,1999).

A mineração de textos explora técnicas e metodologias provenientes das áreas de recuperação de informação, extração de infor-

mação, e linguística computacional de corpus. Para extrair informações úteis, são descobertas características relevantes nos documentos, cujas mais utilizadas são: caracteres, palavras, termos e conceitos. Podem ser realizadas abordagens híbridas para a geração de representações de documentos baseadas em características (FELDMAN; SANGER, 2007). Por exemplo, pode-se primeiro extrair termos do texto, então adaptá-los ou normalizá-los, comparando-os a uma lista de tópicos relevantes (conceitos), extraídos através de uma categorização.

As tecnologias empregadas na mineração incluem extração de informação, rastreamento de tópicos, produção de sumários, categorização de textos, agrupamento de textos, links conceituais, visualização de informação, análise de perguntas e respostas (FAN *et al.*, 2006). Uma técnica bastante comum utilizada é a representação das características de um documento através de um modelo de espaço vetorial. Nesta técnica, cada termo do documento torna-se uma característica dimensional. O valor de cada dimensão pode indicar o número de vezes que o termo aparece no texto, ou pode indicar o peso do termo a ser considerado, como por exemplo, a quantidade de documentos nos quais o termo aparece. No entanto, essa técnica descarta informações importantes, como por exemplo, a ordem nos quais os termos aparecem, onde os mesmos aparecem, e a proximidade entre eles (SCHENKER, 2003).

Os grafos são construções matemáticas importantes e efetivas para realizar a modelagem de relacionamentos e de informação estrutural. Os grafos são utilizados em diversos tipos de problemas, incluindo ordenação, compressão, análise de tráfego, alocação de recursos, entre outros. Como os grafos retêm mais informações do que vetores de características atômicas simples, eles repre-

sentam uma modelagem valiosa, e podem ser utilizados para a representação de textos (SCHENKER, 2003).

A técnica de mineração de textos utilizando grafos descobre as palavras com maior ocorrência no texto, e identifica se elas estão próximas. O grafo obtido a partir da mineração apresenta em seus nós as palavras que mais ocorreram. As associações entre os nós do grafo indicam a proximidade entre as palavras.

O *software* Sobek (LORENZATTI, 2007), cuja interface inicial com um texto a ser minerado é apresentado na Figura 1, é um exemplo de utilização de mineração de textos utilizan-

do grafos. Esse software permite construir um grafo a partir de informações estatísticas obtidas do texto, onde os vértices e as arestas possuem informações referentes ao número absoluto e relativo de ocorrência dos termos (vértices) e associações (arestas) em determinado documento. O grafo obtido a partir do documento representa uma rede dos conceitos que foram trabalhados dentro do texto.

No Sobek, a primeira etapa de processamento compreende a análise léxica, onde o texto é desmembrado palavra por palavra. A seguir, elas são submetidas à análise estatística. Nesse momento, com base nos dados estatísticos, é

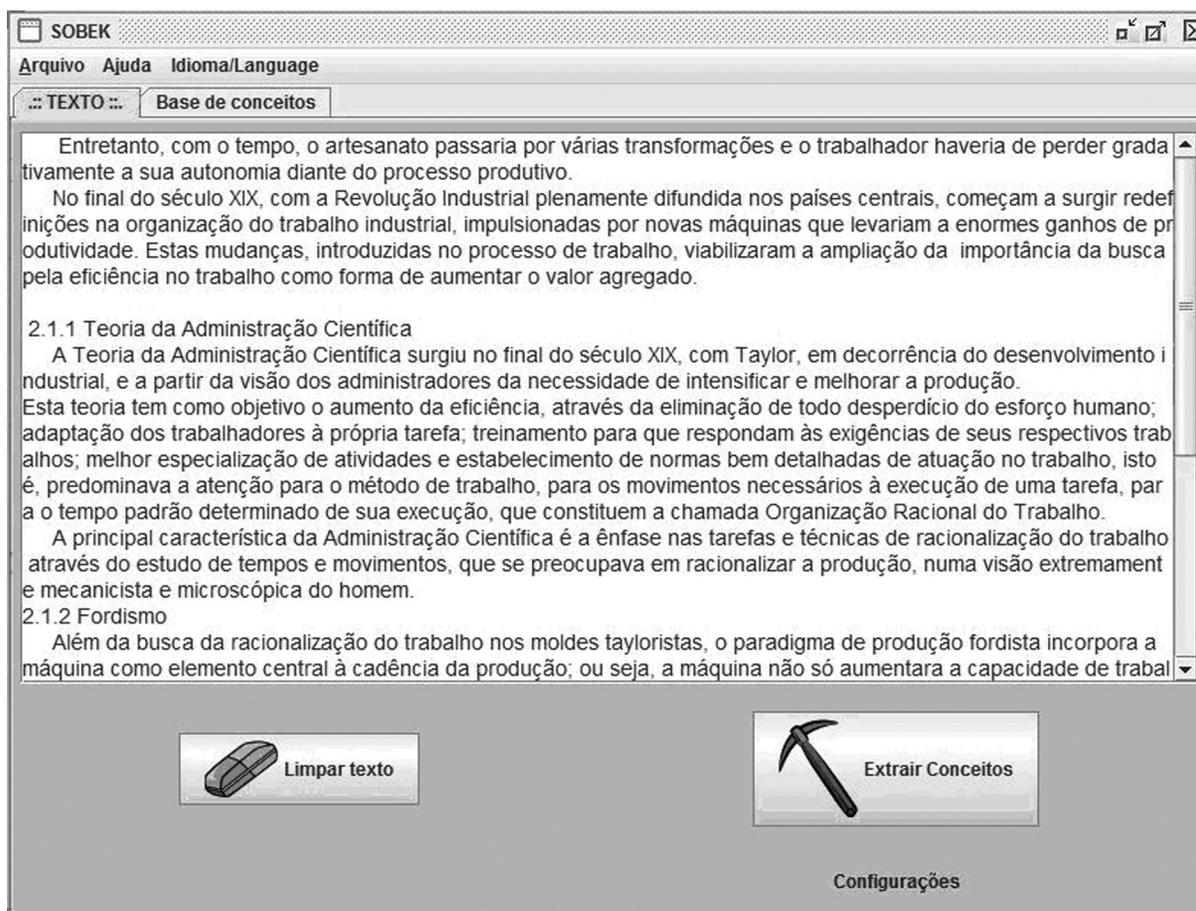


FIGURA 1 – Interface do Sobek.
FONTE: Lorenzatti, 2007, p. 27

criada a base de conceitos. Os conceitos podem ser compostos de uma ou mais palavras. O passo seguinte consiste na remoção de palavras que não agregam significado ao texto, tais como: artigos, conjugações dos verbos ser, estar, ter e haver, bem como de pronomes.

Feita a remoção das palavras e em posse da análise estatística, cria-se um grafo. Nessa atividade, parâmetros como frequência absoluta e tamanho do conceito (se simples ou composto) são analisados. A interface do Sobek permite selecionar o tamanho do grafo desejado. Dentre as opções tem-se: pequeno, médio, grande e enorme. Os tamanhos dos grafos obedecem a perfis de configuração, uma vez que os mesmos contêm os parâmetros utilizados na mineração dos textos. Tais parâmetros definem, por exemplo, o número mínimo de ocorrências que um termo deve ter para integrar-se ao grafo.

Destaca-se que no Sobek os vértices do grafo podem possuir mais de uma palavra. Assim, criou-se uma métrica que faz a seleção dos conceitos e conseqüentemente dos vértices. No momento da seleção, a semelhança entre os conceitos é avaliada, bem como o tamanho dos mesmos (número de palavras) e a frequência absoluta. Para isso, o cálculo é realizado entre conceitos semelhantes (composto de mesma palavra), priorizando os maiores, que expressam ideias mais abrangentes (LORENZATTI, 2007). Destaca-se que o Sobek foi utilizado neste trabalho como a base de desenvolvimento da Rede de Conceitos.

4 Rede de Conceitos

A Rede de Conceitos é uma ferramenta desenvolvida com base na tecnologia de mineração de texto e tem por objetivo extrair e interligar os principais termos tratados em uma produção textual. O intuito é que o resultado

retornado seja suficiente para permitir que o usuário tenha noções sobre o tema abordado e os desdobramentos contemplados na produção, mesmo sem ter lido o texto.

A Rede foi implementada a partir do *software* Sobek, apresentado na seção anterior, e integrada ao ETC, permitindo que os textos desenvolvidos nele sejam facilmente minerados. Sua aplicação é recomendada para todo tipo de público que venha a desenvolver práticas no Editor e que queira visualizar e entender a perspectiva abordada no texto, sem ter que necessariamente ler todo o texto na íntegra.

4.1 A Aplicação da Rede de Conceitos em Textos Coletivos

Macedo et al. (2009) e Macedo (2010), propõe a utilização da Rede de Conceitos no Editor de Texto Coletivo – ETC, disponível em <http://www.nuted.ufrgs.br/etc2>, com o objetivo de apoiar o trabalho do professor, permitindo identificar através da visualização dos grafos, problemas nos textos envolvendo coesão, coerência, entre outros.

A Rede de Conceitos destacada na Figura 2 foi extraída de um texto produzido em um curso de extensão e apresenta os conceitos que aparecem com maior frequência.

Considerando esta Rede de Conceitos, mesmo sem ter lido o texto de origem, é possível perceber que a mesma aborda eixos como: equipes de trabalho, competência e tecnologias digitais. Tal observação pode ser feita se forem destacados, por exemplo, apenas dez dos termos exibidos na Rede da Figura 2, dos quais: equipes de trabalho, colaborativa, social, interação, competências, organizações, flexibilidade, comunicação TICs, recursos e internet.

Para relacionar o resultado dessa Rede de Conceitos com o texto que a originou e visando sustentar a argumentação feita, a seguir são exibidos alguns extratos que correspon-

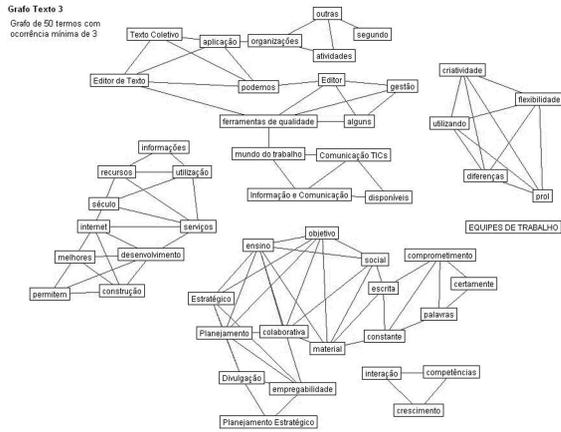


FIGURA 2 – Rede de Conceitos Texto 3 – Grupo: Trabalho em Equipe
FONTE: Macedo, 2010, p. 114

dem, pelo menos em parte, aos termos destacados com o objetivo de evidenciar o assunto que foi abordado. Os extratos estão organizados na Quadro 1 por parágrafos, completos ou não. Cada linha corresponde a um parágrafo, o qual não necessariamente está em sequência com a linha seguinte no texto original.

<p>As empresas necessitam de colaboradores que saibam trabalhar em equipe, utilizando a criatividade, a flexibilidade e a cooperação em prol de objetivos comuns. Focalizamos, no primeiro item, as diferenças entre grupos e equipes e como obter sucesso ao trabalhar em conjunto, respeitando a individualidade de cada membro das mesmas.</p>
<p>No mundo atual onde as informações são produzidas e consumidas de forma virtual, instantânea, em grande quantidade, com abrangência global, a utilização intensa de recursos que aproximem pessoas, divulguem ideias e economizem tempo e energia são fundamentais. As novas Tecnologias da Informação e Comunicação (TICs), as ferramentas de gestão bem como os novos conceitos de trabalho colaborativo – em equipes reais ou virtuais, já são uma realidade neste novo século.</p>
<p>O uso cada vez mais abrangente das Tecnologias da Informação e Comunicação (TICs) trouxe, na nova era digital, a possibilidade das produções escritas decorrentes dos trabalhos em equipe serem feitas de forma assíncrona, ou seja, em tempos e locais diversos, com mais rapidez e objetividade. Há flexibilização na maneira de produzir os textos colaborativos sem que o comprometimento, a interação e a criatividade se percam. Esta flexibilidade e rapidez alicerçam a busca por maior competitividade.</p>

QUADRO 1 – Extratos do Texto 3 – Grupo: Trabalho em Equipe
FONTE: Macedo, 2010, p. 115

Os extratos do texto permitem perceber o contexto em torno da produção escrita e remetem à ideia dos termos destacados inicialmente como indicadores do eixo temático.

Ressalta-se que a análise feita na íntegra², tanto da Rede de Conceitos quanto do texto, fornece melhores instrumentos de análise e conclusões. Neste estudo, para fins de organização e objetividade, buscaram-se fragmentos indicativos das possibilidades encontradas a partir da leitura e interpretação geral dos elementos coletados.

Além da identificação ou não do tema, notou-se que a estrutura da Rede tem relação direta com a qualidade da produção textual, isto é, as Redes, de certa forma, 'denunciam', por exemplo, textos com necessidade de aprofundamento teórico ou com justaposição de parágrafos. Um exemplo de necessidade de aprimoramento pode ser vista na Rede representada pela Figura 3.

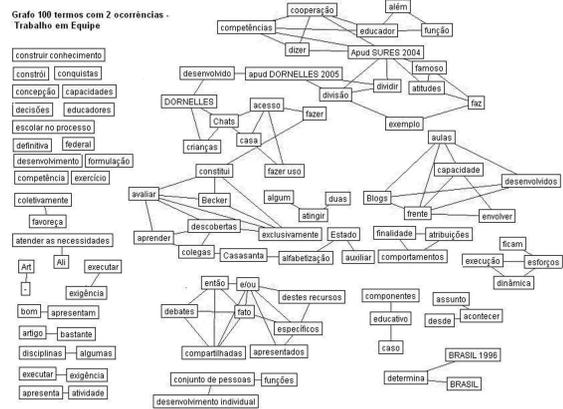


FIGURA 3 – Rede de Conceitos Texto 4 – Grupo: Trabalho em Equipe
FONTE: Macedo, 2010, p. 122

Nessa Rede de Conceitos, nota-se a incidência de vários termos soltos, isto é, termos que não estão relacionados com nenhum outro

2 Uma abordagem detalhada sobre os dados destacados pode ser feita em Macedo (2010).

da Rede e que por si só não oferecem suporte suficiente para entender o contexto em que o termo foi abordado. Além disso, também foi identificada a incidência de vários pequenos conjuntos de termos. Nesse caso, é possível perceber que a identificação do tema central, bem como de seus possíveis desdobramentos, fica comprometida. Os conceitos isolados são indicadores bastante vagos, pois não oferecem condições para conhecer o cenário da abordagem textual. Depois da análise da Rede em questão, fez-se a leitura do texto que a originou onde se verificou que os inúmeros conceitos que estão soltos na Rede, revelam um texto que também pode ser considerado 'solto' na sua essência. Solto porque cita vários diferentes assuntos, mas não trata de nenhum com real propriedade, tampouco relaciona-os na totalidade da escrita. É um texto que necessita de aprimoramento em relação à coerência e coesão. Assim, pode-se entender que se trata de um texto com necessidade de aprimoramento.

De outro lado, a análise das Redes apoiou-se também nas que apresentaram incidência de termos conectados entre si, como representado na Figura 4.

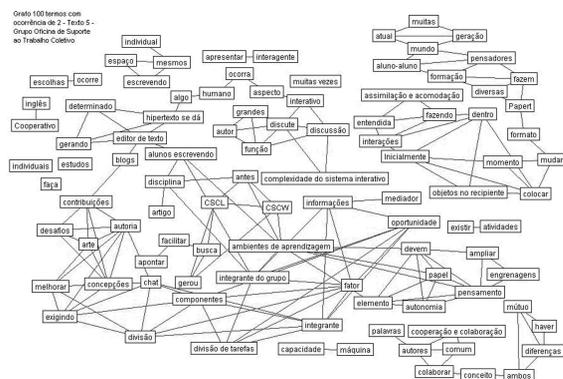


FIGURA 4 – Rede de Conceitos texto 5 – Oficinas Virtuais de Aprendizagem
FONTE: Macedo, 2010, p. 133

Nesse caso, nota-se que prevalecem conjuntos de conceitos relacionados entre si, contrapondo-se à incidência de termos isolados do exemplo. A interpretação da Rede de Conceitos, nessa situação, indica uma produção que trata da escrita coletiva a distância. Destaca-se que os conceitos, de forma geral, se complementam formando uma unidade temática. Por exemplo: editor de texto, ambientes de aprendizagem, *blog*, *chat*, divisão de tarefas, cooperação e colaboração, complexidade do sistema interativo, discussão, contribuições, autoria e desafios, são alguns dos conceitos que, mesmo não estando inteiramente interligados por arestas, complementam-se na indicação do tema central da produção textual.

Feita a análise da Rede de Conceitos, partiu-se para a leitura do texto, onde a introdução aponta os eixos que serão contemplados na produção textual, entre eles, o processo desenvolvido pelo grupo na atividade de EAD e conceitos como CSCL, CSCW, cooperação, colaboração, interação e aprendizagem, além de discussão sobre tecnologias na educação. A primeira seção trata da conceitualização dos termos CSCW, CSCL, cooperação e colaboração. Estes dois últimos conceitos introduzem o termo interação, que é definido, fundamentado e contextualizado, e apresentando em subseções eixos como: interação e hipertexto, interação e aprendizagem. Essa abordagem introduz o tópico que fala sobre interação e a escrita coletiva computacional e segue com foco no uso do ETC (Editor de Texto Coletivo) em relação ao processo do grupo e suas aplicações. A produção textual finaliza com considerações sobre o uso da tecnologia como possibilidade de apoio à aprendizagem. Destaca-se que esse texto diferencia-se do apresentado no exemplo anterior em termos de estrutura. Aqui, o fluxo do texto apresenta encadeamento, bom desenvolvimento, lógica na abordagem e boa argumentação.

Com base na aplicação e análise da Rede gerada a partir dos textos produzidos no ETC, destaca-se que ela mostrou-se suficientemente eficiente, quando a intenção foi conhecer o tema abordado na produção textual, uma vez que o intuito foi entender do que trata o texto, sem necessariamente ter lido este antes de gerar a Rede. Além disso, destaca-se que a Rede de Conceitos oferece indicadores sobre a produção textual para que o professor possa agir com foco nas necessidades e potencialidades dos alunos, sem necessariamente ter que ler o texto produzido, intenção principal que disparou o interesse por este estudo.

4.2 A Aplicação da Rede de Conceitos nos Fóruns de Discussão

Outra experiência em que foi aplicada a Rede de Conceitos é apresentada por Azevedo *et al.* (2009). Este propõe a utilização da mesma para minerar fóruns de discussão, em ambientes virtuais de aprendizagem. O objetivo foi identificar se as discussões no fórum possuem relevância, ou se tem conexão, integração entre os conceitos relacionados ao tema do debate.

Em um fórum de discussão, dependendo do tema proposto, do perfil dos participantes, do interesse despertado pelo tema, podem-se ter centenas de contribuições. Alguns textos redigidos possuem relevância com relação ao tema tratado, e outros não. Observa-se que as contribuições significativas abordam conceitos importantes relativos ao assunto em debate, onde estes conceitos são citados e relacionados entre si.

Para que o professor possa acompanhar minuciosamente todas as contribuições colocadas pelos alunos em um fórum de discussão é necessário um tempo bastante considerável. Desta forma, é importante proporcionar ao professor uma forma de visualizar as contri-

buições relevantes registradas no fórum a fim de auxiliar o docente nas intervenções destes com os alunos envolvidos.

A utilização da Rede de Conceitos para analisar fóruns de discussão está em fase de desenvolvimento. Por essa razão, neste artigo são apresentados os resultados de um estudo piloto.

Para minerar os textos dos fóruns, foi seguido o mesmo princípio de extração das Redes de Conceitos. Assim, a partir das redes geradas, realizou-se manualmente³ um estudo das mesmas, para analisar quais contribuições foram relevantes de acordo com o tema proposto.

A fim de viabilizar esta atividade, definiu-se o coeficiente de relevância temática – CRT (Azevedo *et al.*, 2009) de uma contribuição textual, cujo objetivo é analisar o quanto um texto é relevante a um determinado tema de discussão. O coeficiente de relevância temática indica o grau de relevância da contribuição textual de acordo com o tema do fórum. O coeficiente de relevância pode ser calculado a partir da seguinte fórmula:

- NC: número de conceitos relevantes utilizados no texto
 - NA: número de associações entre os conceitos relevantes utilizados no texto
- $$CRT = NC + NA$$

Para o cálculo do NC, os termos considerados como equivalentes semânticos também são considerados como conceitos relevantes. Os prefixos, sufixos, e plurais dos conceitos relevantes citados no texto são convertidos, para também serem considerados como conceitos relevantes. Com o CRT, as redes que possuem mais conceitos importantes relativos ao tema, com mais associações entre os con-

³ O processo de mineração dos fóruns está em fase de testes e implementação, por isso diz-se que, neste momento a análise foi feita de forma manual.

ceitos, possuem maior coeficiente de relevância temática.

A metodologia empregada envolveu:

- A escolha de um fórum de discussão.
- A definição dos conceitos importantes relativos ao tema do fórum, bem como a definição da associação entre estes conceitos. Os conceitos podem ser digitados e relacionados pelo professor, ou um texto de referência pode ser indicado pelo docente. Os conceitos e suas relações serão extraídos do texto pela ferramenta.
- A definição de possíveis termos que podem ser considerados como equivalentes semânticos aos conceitos envolvidos.
- A definição do valor mínimo do coeficiente de relevância temática que será considerado para análise.
- A coleta das contribuições textuais produzidas pelos alunos no fórum.
- Para cada contribuição textual, foi gerada uma Rede de Conceitos.
- A partir de cada Rede, foi calculado o coeficiente de relevância temática da contribuição textual.
- A partir da organização realizada com as Redes geradas, as contribuições textuais relativas as mesmas foram organizadas.
- A partir das Redes geradas foi realizada uma análise da quantidade de contribuições relevantes colocadas por cada aluno no fórum.

As mensagens postadas por cada aluno no fórum são consideradas. Cada mensagem é analisada individualmente e a partir do texto redigido é construída uma Rede de Conceitos. Em cada Rede, é realizado o cálculo do coeficiente de relevância temática da contribuição textual. Por exemplo, se em um fórum par-

ticiparam 20 alunos, e cada aluno redigiu 10 mensagens, serão geradas e avaliadas 200 Redes de Conceitos.

O primeiro estudo foi realizado em um fórum de discussão do Ambiente ROODA, na disciplina *Tópicos Especiais Z1* da turma do Doutorado em Informática na Educação, que ocorreu no primeiro semestre de 2008. O tema do fórum foi *Comunidades Virtuais*. O fórum contou com a participação de 05 alunos, que realizaram 25 contribuições textuais.

Para este fórum, o professor definiu como importantes, os seguintes conceitos: *comunidades, virtuais, comunidades virtuais, aprendizagem, comunidades virtuais de aprendizagem, educação, relações sociais*. As associações relevantes entre os conceitos ficaram estabelecidas da seguinte maneira: o conceito *comunidades virtuais de aprendizagem* está relacionado com os conceitos *comunidades, comunidades virtuais, aprendizagem, educação*. Os termos CVs e CVAs foram definidos como equivalentes semânticos aos termos *comunidades virtuais* e *comunidades virtuais de aprendizagem*, respectivamente.

Foi definido o valor mínimo 01 para o coeficiente de relevância temática (CRT) a ser considerado na análise. O coeficiente de relevância temática foi calculado a partir da rede de conceitos gerada em cada contribuição textual. Como um exemplo de cálculo do CRT, pode-se observar a mensagem abaixo extraída do fórum, postada pelo aluno C.

Mensagem: *Acredito que as CVAs podem oferecer um recurso valioso à educação, pois elas possibilitam um mecanismo de construção coletiva de conhecimento, através das discussões realizadas pelos membros da comunidade.*

A rede gerada a partir da mensagem anterior pode ser visualizada na Figura 5.

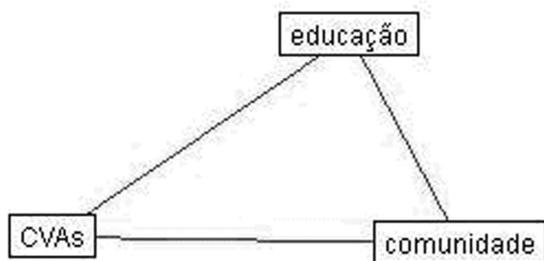


FIGURA 5 – Rede gerada a partir de uma mensagem do fórum do experimento 1.
FONTE: Azevedo, 2009, p. 80

Para a Rede anterior, o CRT é calculado da seguinte maneira:

$$NC = 3$$

$$NA = 3$$

$$CRT = NC + NA = 6$$

Foi realizada uma análise da quantidade de contribuições relevantes colocadas por cada aluno no fórum. Os resultados são apresentados no gráfico da Figura 6.

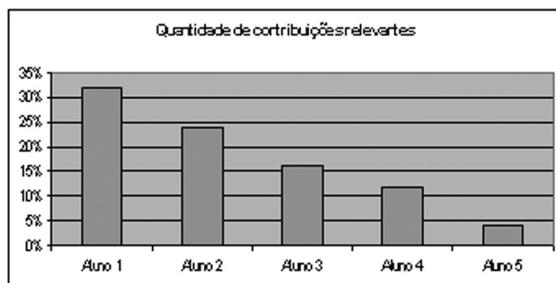


FIGURA 6 – Quantidade de contribuições relevantes do primeiro experimento.
FONTE: o autor

A partir da Figura 6, percebeu-se que a utilização da metodologia proposta permitiu verificar que a maior parte das contribuições feitas pelos estudantes era relevante ao tema em debate. Este estudo considera como relevantes as mensagens postadas que citaram conceitos relacionados ao tópico da discussão,

cujo CRT seja maior que o mínimo estabelecido pelo docente.

Na época que os experimentos foram realizados, a análise qualitativa das mensagens postadas nos fóruns foi pautada no cálculo do CRT de cada uma. Atualmente, um software para mineração de textos está sendo desenvolvido para analisar automaticamente as mensagens de fóruns de discussão. Este software aprimora a fórmula do CRT, para também considerar o peso de cada conceito. Nos experimentos apresentados neste artigo, o cálculo do CRT de cada contribuição textual foi realizado manualmente. No software que está em desenvolvimento, o cálculo é efetuado automaticamente pelo *software*.

O segundo e terceiro experimentos foram realizados em um fórum de discussão do Ambiente ROODA, na disciplina *Seminário Integrador VII – B* de uma turma do Curso de Graduação em Pedagogia, que ocorreu no segundo semestre de 2009. O tema do fórum foi *Aprender com os outros*. A turma foi dividida em dois grupos, e cada grupo contribuiu em um fórum de discussão. O primeiro fórum contou com a participação de 27 alunos, que realizaram 45 contribuições textuais. O segundo fórum contou com a participação de 29 alunos, que realizaram 67 contribuições textuais.

Para estes fóruns, o professor definiu como importantes, os seguintes conceitos: *aprender, aprendizagem, conhecimento, construção do conhecimento, Piaget, afeto, afetividade, estruturas cognitivas, descentração, relações afetivas, experiências afetivas, laços afetivos, cooperação, colaboração*. Foram estabelecidas as associações relevantes entre estes conceitos.

Foi definido o valor mínimo 01 para o coeficiente de relevância temática (CRT) a ser considerado na análise. O coeficiente de relevância temática foi calculado a partir da rede de conceitos gerada em cada contribuição

textual. Foi realizada uma análise da quantidade de contribuições relevantes colocadas por cada aluno do primeiro e segundo grupos. Os resultados são apresentados nas Figuras 7 e 8, respectivamente.

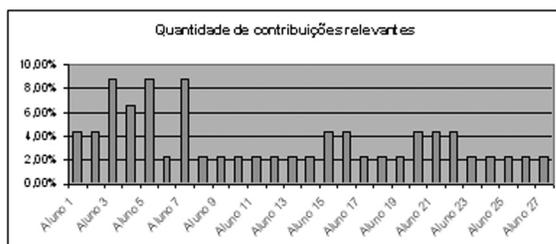


FIGURA 7 – Quantidade de contribuições relevantes do primeiro grupo.
FONTE: o autor

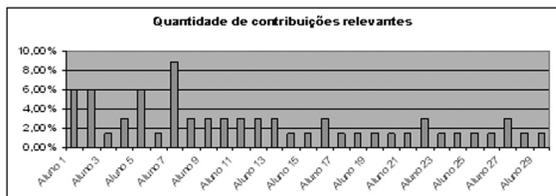


FIGURA 8 – Quantidade de contribuições relevantes do segundo grupo.
FONTE: o autor

A partir dos resultados apresentados nas Figuras 6, 7 e 8, pode-se verificar a quantidade de contribuições relevantes ao tema do fórum produzidas pelos alunos. Percebe-se que a metodologia empregada auxilia o docente a identificar a quantidade de mensagens postadas por cada discente, nas quais foram citados conceitos relacionados ao tópico em debate.

Observou-se que a versão atual da Rede de Conceitos, utilizada nos experimentos pilotos, realiza perfeitamente a mineração de um texto e a geração do grafo representativo dos conceitos encontrados no mesmo. No entanto, as outras atividades desenvolvidas nos experimentos foram executadas manualmente.

Para permitir que a Rede de Conceitos pos-

sa oferecer os recursos necessários ao professor, será necessário aprimorá-la e criar novas funcionalidades como:

1. Realizar o *stemming*⁴ de palavras.
2. Permitir a leitura de um conjunto de conceitos que possuam equivalência semântica.
3. Permitir a identificação de perguntas feitas pelos alunos.
4. Permitir a leitura automática dos textos provenientes de um fórum de discussão.
5. Permitir a definição do valor mínimo do coeficiente de relevância temática que será considerado para análise.
6. Executar o cálculo do coeficiente de relevância temática para cada grafo gerado a partir dos textos produzidos pelos alunos.
7. A partir dos grafos gerados, realizar uma análise da quantidade de contribuições relevantes colocadas por cada aluno no fórum.
8. Gerar um relatório visual para o professor, contendo informações sobre as contribuições relevantes de um tema de discussão. Estas informações disponibilizadas ao professor indicarão a quantidade de contribuições relevantes de cada aluno e a quantidade de contribuições não relevantes.

Com base nos resultados deste trabalho, verificou-se que a metodologia apresentada pode auxiliar o professor a realizar diversas atividades relacionadas ao fórum, como:

- Analisar quais contribuições textuais necessitam de intervenção.
- Visualizar os alunos que colocaram apenas algumas contribuições relevan-

⁴ *Stemming* é a redução de um conjunto de palavras ao mesmo *stem*, através da retirada dos prefixos e sufixos, permanecendo apenas a raiz das palavras (SPARK-JONES; WILLET, 1997).

tes no fórum para oferecer-lhes maior auxílio.

- Estimular os alunos que colocaram muitas contribuições relevantes para interagir com os alunos que colocaram poucas.

A partir das atividades citadas, o professor pode direcionar seu auxílio aos alunos que registraram poucas contribuições relevantes. O professor pode também, estimular um processo de aprendizagem colaborativa, sugerindo a interação entre os alunos que mais registraram textos relevantes com os alunos que redigiram poucos.

5 Considerações Finais

No que tange à aplicação de recursos de mineração texto na produção textual coletiva, destaca-se que a Rede de Conceitos fornece informações sobre o tema desenvolvido na produção textual e indicadores qualitativos da produção. Os indicadores qualitativos diferenciam os textos que necessitam de aprimoramento dos que foram desenvolvidos com clareza e objetividade. Todas essas situações podem ser identificadas através da Rede, sem a leitura prévia do texto que as originou. Nes-

sas condições, o tempo, que antes era dedicado à leitura desse material, agora pode ser investido na intervenção junto aos alunos, a partir de diferentes práticas pedagógicas, com o objetivo de potencializar o processo de aprendizagem e qualificar a produção escrita.

Em relação às discussões realizadas nos fóruns de discussão, entende-se que a Rede de Conceitos pode ser utilizada na criação de indicadores de relevância das postagens de um fórum, estabelecendo assim, relações com os textos desenvolvidos pelos alunos e, a partir daí, verificar a pertinência e o processo de desenvolvimento e construção de cada produção. Destaca-se que todo esse envolvimento pode ser feito em um tempo menor que o exigido atualmente, em função dos recursos disponibilizados pela técnica de mineração de textos.

É importante observar que a Rede de Conceitos está sendo aprimorada para ser adaptada a diferentes Ambientes Virtuais de Aprendizagem e, inclusive para trabalhar com vários formatos de editores de texto e fóruns de discussão. Outro passo interessante e que está em processo de análise é a implementação da Rede de Conceitos como uma API que possa ser adaptada a AVAs com tecnologias variadas, como PHP e Python.

Referências

- ASCOTT, R. Cultivando o Hipercórtex. In: DOMINGUES, D. (Org.). *A Arte no Século XXI: a humanização das tecnologias*. São Paulo: Fundação Ed. da UNESP, 1997. P. 336-344.
- AZEVEDO, B.F.T. *et al.* Estudo de Análise Qualitativa em Fórum de Discussão. *RENOTE: Revista Novas Tecnologias na Educação*, Porto Alegre, v. 7, n. 2, p. 76-86, 2009.
- FAN, W. *et al.* Tapping the power of text mining. *Communications of ACM*. Nova York, v. 9, n. 49, p. 76-82, 2006.
- FELDMAN, R.; SANGER, J. *Text Mining Handbook*. Inglaterra: Universidade de Cambridge, 2007.

HARASIM, L.; CALVERT, T.; GROENEBOER, C. *Virtual-U: A Web-based System to Support Collaborative Learning*. Englewood Cliffs, NJ: Educational Technology Publications, 1997.

HEARST, M.A. *Untangling Text Data*. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 37., 1999, Maryland. *Proceedings of ACL'99*. Stroudsburg, PA: Association for Computational Linguistics, 1999.

LÉVY, P. *As Tecnologias da Inteligência*. Rio de Janeiro, 1993.

LORENZATTI, A. *SOBEK: uma Ferramenta de Mineração de Textos*. 2007. Trabalho de Conclusão – Departamento de Informática, Universidade de Caxias do Sul, 2007, Caxias do Sul, BR-RS.

MACEDO, A. *Rede de Conceitos: uma ferramenta para contribuir com a prática pedagógica no acompanhamento da produção textual coletiva*. 2010. Tese (Doutorado) – Programa de Pós-Graduação em Informática na Educação, Universidade Federal do Rio Grande do Sul, 2010, Porto Alegre, BR-RS.

MACEDO, A. et al. P. Using Text-Mining to Support the Evaluation of Texts Produced Collaboratively. In: TATNALL, A.; JONES, A. (Org.). *Education and Technology for a Better World: Selected papers of the 9th IFIP TC3 World Conference on Computers in Education, WCCE*. Berlim: Springer, 2009. P. 368-377.

SCHENKER, A. *Graph-Theoretic Techniques for Web Content Mining*. 2003. Tese (Doutorado em Ciência da Computação) – University of South Florida, 2003, Estados Unidos.

SPARK-JONES, K.; WILLET, P. *Readings in Information Retrieval*. San Francisco: Morgan Kaufmann, 1997.

TAN, A. Text mining: the state of the art and the challenges. In: PACIFIC-ASIA WORKSHOP ON KNOWLEDGE DISCOVERY FROM ADVANCED DATABASES – PAKDD'99, 1999, Beijing. *Proceedings*. [S.l.]: Springer, 1999. P. 65-70.

*Recebido em 15 de maio de 2010.
Aprovado para publicação em 03 de março de 2011.*

Alexandra Lorandi Macedo

Universidade Federal do Rio Grande do Sul (UFRGS). Porto Alegre/RS – Brasil. E-mail: alorandimacedo@gmail.com

Breno Fabrício Terra Azevedo

Universidade Federal do Rio Grande do Sul (UFRGS). Porto Alegre/RS – Brasil. E-mail: brenoter@gmail.com

Patricia Alejandra Behar

Programa de Pós-Graduação em Informática na Educação (PPGIE) da Universidade Federal do Rio Grande do Sul (UFRGS). Porto Alegre/RS – Brasil. E-mail: pbehar@terra.com.br

Eliseo Reategui

Programa de Pós-Graduação em Informática na Educação (PPGIE) da Universidade Federal do Rio Grande do Sul (UFRGS). Porto Alegre/RS – Brasil. E-mail: eliseoreategui@gmail.com