

# Aplicação Educacional de uma Ferramenta de Mineração de Textos Integrada a uma Ontologia de Domínio na Área da Saúde

Fábio Rafael Damasceno – PGIE – UFRGS – fabiorafaeldamasceno@gmail.com

Alexandre Ribeiro – DEIN – UCS – aribeiro@ucs.br

Eliseo Reategui – PGIE/PPGEDU – UFRGS - eliseoreategui@gmail.com

**Abstract.** *This article shows the educational usage of a textual mining tool (Sobek) inside the Communications and Information Technology Observatory of Health Science (CITOHS). With the main goal of giving users new possibilities for exploring the Observatory, we investigated how a domain ontology could be associated with its International Classification of Diseases (ICD). This study shows how the integration of the mining tool with this domain ontology, together with the ICD categorization, may relate documents from a semantic point of view, enabling a more dynamic search, browsing and access to these documents.*

**Keywords:** *Text-mining, ontologies*

**Resumo.** *Este artigo aborda a utilização educacional de uma ferramenta de mineração textual (Sobek) dentro do Observatório de Tecnologias de Informação e Comunicação das Ciências da Saúde (OTICS). Com o objetivo de tornar a exploração em tal ambiente diferenciada, foi investigada a possibilidade de associar uma ontologia de domínio na área da saúde à sua Classificação Internacional de Doenças (CID). O estudo mostra como a integração da ferramenta de mineração com esta ontologia de domínio, junto da catalogação CID, pode relacionar semanticamente documentos, possibilitando uma busca, navegação e acesso mais ágil.*

**Palavras-chave:** *Mineração textual, ontologias*

## 1. Introdução

Entre os avanços obtidos pelo Sistema Único de Saúde (SUS) no país nos últimos anos está a aprovação da política para a área de informação e comunicação, abordando aspectos críticos apontados por estudos dos órgãos de gestão e participação no sistema de saúde. Esta política associa o uso da informação propriamente dito para apoio à decisão em diferentes níveis do sistema de saúde, com o uso criativo e inovador dos recursos informacionais e de comunicação. Esse foi o intuito da primeira fase do Projeto Observatório de Tecnologias de Informação e Comunicação em Sistemas e Serviços de Saúde (Projeto OTICS).

O Projeto OTICS busca desenvolver um observatório capaz de integrar e articular os Sistemas Nacionais de Informação em Saúde do Ministério da Saúde e os Bancos de Dados Nacionais oriundos dos inquéritos e levantamentos censitários

realizados pelo IBGE, contribuindo para a ampliação da inteligência coletiva sobre a saúde e o sistema de saúde no país.

Ambientes de colaboração na Web, como é o caso do Projeto OTICS, possuem um caráter interdisciplinar e público bastante heterogêneo, gerando e requisitando conteúdos de variadas categorias. Fóruns, *chats* e repositórios de documentos são as ferramentas que facilitam o acesso e a troca de informações entre os atores envolvidos nestes ambientes. No entanto estes recursos não são relacionados levando em consideração o seu significado, o seu real sentido. Comumente estão agrupados em módulos fixos, organizados em contribuições datadas semanalmente por determinados grupos de discussão. Considerar estes materiais sob o ponto de vista semântico, através de uma camada adicional de interrelação, poderia possibilitar a exploração destes conteúdos de maneira mais fácil, associando-os a outras bases de documentos.

O cenário descrito acima retrata uma lacuna ainda não preenchida nestes ambientes. As relações semânticas comentadas previamente poderiam contribuir para a construção de ambientes de caráter dinâmico e adaptável na apresentação, criação e exploração de novos conteúdos. Ontologias, estruturas capazes de modelar domínios e suas relações, podem contribuir neste meio para a modelagem semântica destes materiais, viabilizando formas mais ricas de interação.

A adaptação ao aluno ou usuário de um ambiente é retratada desde a década de 1950 por Bloom (1984), que afirma que o ensino individualizado é considerado o melhor caso para o desenvolvimento de um aluno. Esta interpretação poderia ser aplicada aos processos de aprendizagem possibilitados pela exploração de conteúdos em um portal, como o OTICS. Os conceitos relacionados à aprendizagem individualizada impulsionaram a criação dos IACs (Instrução Assistida por Computador) e posteriormente dos STIs (Sistemas Tutores Inteligentes), ambientes que adotaram um modelo humano como tutor em sua estrutura, para proporcionar tal modalidade de ensino/aprendizagem (MURRAY, 1999).

A grande quantidade de documentos que compõe ambientes como o Portal OTICS torna a tarefa de interpretação e extração de conhecimento a partir de fontes textuais algo não-trivial (SHARP, 2001; SOY, 2003). A tentativa de automatizar este processo com auxílio computacional fez surgir a área de pesquisa conhecida como mineração textual. A ferramenta Sobek (PINHO et al., 2011) é uma ferramenta desta categoria que estatisticamente extrai termos relevantes de um texto, representando-os em um grafo, cujo resultado poderia auxiliar no enquadramento de conteúdos para tais ambientes. No entanto, a representação em forma de mapa gerada pela ferramenta carece de semântica, vista a abordagem estatística de extração destas informações, situação em que as ontologias poderiam contribuir significativamente.

Duas linhas de pesquisa neste grande cenário descrito nos parágrafos anteriores estão em crescente expansão: a primeira delas é a Web Semântica, que desenvolve tecnologias que permitem ao computador compartilhar e manipular as informações contidas na Web de forma adequada e inteligente, tratando-se de um ambiente fortemente impulsionado pelas pesquisas envolvendo ontologias. Com a utilização desta abordagem, tanto ambientes de aprendizagem como sistemas computacionais podem interagir entre si, trocando informações e auxiliando professores na seleção, combinação e classificação do conteúdo disponível na Web. A segunda área em expansão é a Web 2.0, ou Web Social, em que usuários podem compartilhar e construir conhecimento de forma simples, interativa e colaborativa (ISOTANI e BITTENCOURT, 2009).

Este artigo propõe uma aplicação educacional de uma ferramenta de mineração textual em um ambiente virtual de exploração de conteúdos, trazendo como diferencial o trabalho conjunto com uma ontologia, acarretando em uma característica semântica no processo. Este conjunto resulta na interação diferenciada para o usuário, visto que a exploração de materiais seguirá este tipo de relacionamento semântico.

O artigo está estruturado da seguinte forma: Na seção dois é abordada a Web semântica e Ontologias, enfatizando conceitos importantes para a pesquisa abordada neste trabalho. Na seção três os conceitos de Mineração textual são vistos, servindo de base para a apresentação da ferramenta Sobek na seção seguinte. Já na seção cinco o Portal OTICS é conceituado enfatizando o seu objetivo e possibilidades de aplicação real no cenário de saúde brasileira. Por fim, na seção seis, é apresentada a proposta de integração da ferramenta Sobek a uma ontologia de domínio com vistas ao enriquecimento semântico dos grafos gerados como resultado dos processos de mineração. Tal solução tem como objetivo proporcionar aos usuários do Portal OTICS uma forma dinâmica e interativa de explorar os conteúdos disponibilizados no portal.

## 2. Ontologias e Web Semântica

A *web* está evoluindo de um grande espaço de informação e comunicação para um repositório massivo de conhecimento e serviços. Um conceito que permite esta mudança são as ontologias, estruturas criadas a partir de especialistas em um domínio que contém as relações entre os objetos do domínio em questão (ZHOU, 2007). Atualmente a *Web* está estruturada para lidar com pessoas, termos de domínio e ainda *tags* HTML que são claramente compreensíveis por desenvolvedores e usuários, no entanto não possuem sentido algum para sistemas computadorizados (JACOB, 2003).

As interligações na Web são inúmeras, relacionando páginas e arquivos através dos conceitos de hipertexto. Embora sejam de caráter muito amplo, estas ligações parecem não tirar proveito do potencial contido nos significados de suas páginas, documentos e mídias. O que compõe atualmente estas ligações são relações entre os atributos destes artefatos presentes na *Web*. Nomes de arquivo, tipo de extensão e tamanho em bytes são exemplos destas variáveis de interligação - não contemplando o real significado do artefato digital. Estas informações deveriam constituir uma das variáveis de tais inter-relacionamentos, e não o todo como acontece atualmente. Como alternativa para modelagem destas relações e conceitos nesta massa de informações, existe o conceito de Ontologias mencionado previamente (JACOB, 2003).

Este cenário, em que os reais significados de artefatos digitais relacionam objetos digitais, permite a definição da chamada *Web Semântica* (LEE, HENDLER e LASSILA, 2001), que constitui este paradigma de uma Web interligada através dos significados e sentidos destas informações e arquivos.

A informação na Web é representada em linguagem natural, sendo compreensível por pessoas. No entanto, para prover informação de maneira que computadores ou sistemas computacionais possam extrair e representar seu significado, é necessário disponibilizar a informação formalmente e de maneira sistemática (ISOTANI e BITTENCOURT, 2009). Para se tornar realidade, a Web Semântica exige uma estrutura condizente, cuja semântica possa ser interpretada tanto por computadores quanto por humanos – neste cenário as Ontologias ganharam destaque (GRUBER, 1995; GUARINO, 1995). Conhecidas pela possibilidade de modelar diferentes domínios e suas relações, ao mesmo tempo possuindo uma estrutura visual simples

como a de um Mapa Conceitual, as Ontologias fizeram seu espaço e se tornaram a força motriz da nova *Web* que está se formando. Estas oferecem uma linguagem expressiva e formal para gerar informação que possa ser interpretada por computadores (MIZOGUCHI, HAYASHI e BOURDEAU, 2007), permitindo a serem combinadas, compartilhadas, modificadas e utilizadas para definir semanticamente diferentes categorias de recursos (JOVANOVIC *et al.*, 2008). Segundo Devedzic (2006), o uso de ontologias e o desenvolvimento de serviços Web para processar a informação disponível na Internet está transformando a Web da informação (Web tradicional) na Web do conhecimento.

Ontologias fornecem um embasamento semântico sobre descrições de conteúdo digital representáveis por computador, sendo considerada ubíqua em sistemas de informação através de documentos com meta-dados, aprimorando a performance de extração de informação e inferência, tornando dados entre diferentes aplicações interoperáveis. Uma ontologia é construída para facilitar o compartilhamento de conhecimento e reutilização do mesmo entre pessoas. Abstraindo, as linguagens de representação de ontologias formam um espectro que varia entre a linguagem natural em um extremo até a linguagem formal no outro. Estes extremos se aplicam melhor a pessoas e sistemas computacionais, respectivamente. Para diminuir a lacuna entre eles, pode-se permitir que pessoas acessem e modifiquem ontologias utilizando linguagem natural, mapeando esta em uma série de conceitos de baixo nível para que o processo fique automatizado. Isto envolve e aprimora o *aprendizado em ontologias*, processo em que técnicas de aprendizado de máquina são utilizadas para extração de conhecimento (ZHOU, 2007).

Para ser efetivo, um processo de aprendizado, aconteça ele em sala de aula, em um sistema tutor inteligente ou em um ambiente como o OTICS, deve ser adaptável para o contexto do estudante, cenário em que as ontologias podem contribuir, removendo ambiguidades, também ajudando na identificação de categorias semânticas de um domínio (EYHARABIDE, 2009).

### 3. Mineração Textual

A mineração textual tem suas origens mais remotas por volta de 1960 quando Hans Luhn e Lauren Doyle perceberam que a frequência e distribuição de palavras relevantes dentro de um texto eram informações notórias para sua compreensão (SHARP, 2001; SOY, 2003). Nos dias de hoje, as pesquisas em mineração de texto já produziram muitos resultados significativos em diferentes áreas, como apresentado a seguir.

Dalmolin *et al.* (2009) demonstraram uma ferramenta para auxiliar no processo de elaboração de mapas conceituais. A ferramenta foi desenvolvida para organizar conteúdos a partir dos módulos temáticos de um curso, aplicando técnicas desta área de pesquisa. Na ferramenta, os professores podem construir seus próprios mapas baseados em documentos textuais, utilizando um editor visual para vinculá-los a objetos de aprendizagem a conceitos. Os alunos podem acessar objetos de aprendizagem e fazer suas próprias anotações em seus mapas conceituais, melhorando o processo de ensino-aprendizagem. O extrator de mapa conceitual é um algoritmo baseado em técnicas de mineração de texto, extraindo os termos relevantes que podem ser considerados conceitos ou links, tornando a construção do mapa conceitual um processo ágil.

Ferreira (2009) apresenta uma pesquisa sobre a realização da sumarização de textos provenientes de repositórios de arquivos textuais e blogs. De acordo com os autores, os sumários auxiliam as pessoas a identificar mais rapidamente, se o conteúdo do texto original é útil aos seus interesses. O trabalho cita que a sumarização de textos é uma alternativa viável a ser incorporada em Ambientes Educacionais na Web, pois pode ajudar alunos e professores a diagnosticar, com maior velocidade, textos que podem atender as suas necessidades.

Barbosa et al. (2009) apresentam um estudo comparativo entre duas ferramentas para mineração de textos, Sobek e *Tag Clouds*. A mineração foi executada em textos selecionados em blogs. Uma das contribuições do artigo, baseando-se em um contexto educacional, foi que o Sobek possui uma capacidade maior para extrair informações relevantes, uma vez que relaciona termos e exhibe resultados mais completos.

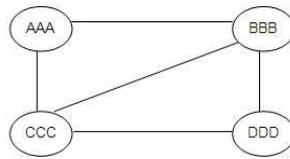
Os trabalhos apresentados nos parágrafos anteriores abordam alternativas para mineração textual, em dados contextos de aplicação. O Projeto Sobek, descrito na seção seguinte, propõe uma ferramenta em que a mineração textual realizada em um documento gera um grafo que relaciona termos encontrados no texto. A pesquisa sendo desenvolvida neste artigo visa o enriquecimento semântico deste processo, através da utilização de uma ontologia de domínio específica para a área da saúde.

#### 4. Sobek

A ferramenta Sobek utiliza o processo de mineração de texto como um método de extração de informações relevantes em bases de dados não estruturadas, ou semi-estruturadas. Tendo suas origens na mineração de dados, a mineração de textos tem se tornado cada vez mais popular, principalmente em função do crescimento da internet e da evolução da área de linguística computacional (SHARP, 2001). A técnica se diferencia da mineração de dados pelo fato de que a última busca por padrões de informação em registros formais de bancos de dados, enquanto a mineração de textos emprega fontes de dados não estruturadas (WITTEN e FRANK, 2005).

A ferramenta Sobek, desenvolvida a partir da técnica de mineração desenvolvida por Schenker (2003), foi inicialmente proposta como mecanismo de apoio ao trabalho docente no acompanhamento da escrita colaborativa (MACEDO et al., 2009). Também foi empregada na avaliação qualitativa das postagens dos estudantes em fóruns de discussão (AZEVEDO, REATEGUI e BEHAR, 2010).

No caso da ferramenta Sobek, um algoritmo específico definido por Schenker (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*). O modelo utiliza um parâmetro  $n$  que faz com que sejam considerados  $n$  termos subsequentes ao termo analisado. Dessa forma o grafo resultante mostra um termo conectado aos  $n$  termos frequentes que aparecem a sua frente. Por exemplo, para  $n=2$ , o texto “AAA BBB CCC DDD” produziria o grafo da Figura 1. Note que não há ligação entre AAA e DDD, sendo a distância do termo “AAA” para “DDD” de 3 (para  $n=2$ , o valor 3 ultrapassa a máxima distância considerada).



**Figura 1: Grafo estatístico gerado pelo Sobek**

Do ponto de vista pedagógico, uma ferramenta de identificação de termos como esta pode permitir aos alunos explorar quais termos e relacionamentos são observáveis em textos de outros autores, ou em suas próprias produções. Um procedimento interativo empregando a ferramenta também foi proposto como forma de apoiar o aluno na realização de produções textuais (KLEMMANN, LORENZATTI e REATEGUI, 2009). Tal atividade poderia ser comparada ao processo de construção textual apoiado pela criação de mapas conceituais. Esses são definidos como diagramas que indicam relações entre termos e palavras usados para representar conceitos (NOVAK e GOWIN, 1984).

## 5. Projeto OTICS

Produzir e socializar o conhecimento adquire uma nova dimensão com o uso da tecnologia. Poucos trabalhos são encontrados que forneçam ferramentas de Gestão de Conhecimento de uso prático e direcionado ao contexto acadêmico. Agregar uma ferramenta de gestão da atividade colaborativa de modo que as próprias práticas (administrativas e pedagógicas) sejam redimensionadas em um contexto de evolução e detalhamento colaborativo, constitui um atrativo fundamental na socialização do conhecimento.

Avaliando cursos da área da saúde, em particular os de medicina, percebe-se que estes utilizam nos ambientes educacionais softwares específicos para algumas disciplinas, como softwares de simulação, de apoio estatístico, de definição de hipóteses diagnósticas ou de revisão de conteúdos de um determinado assunto. Neste novo contexto, os alunos estarão constantemente em contato com a rotina das UBS (Unidade Básica de Saúde). As UBS agrupam postos de saúde que dão todo o tipo de atendimento especializado à comunidade, possuindo formulários e documentos que registram o histórico dos pacientes e os tratamentos conferidos. Estas informações compõem uma preciosa fonte de conhecimento neste novo contexto de aprendizagem, podendo estar disponíveis em um ambiente como o Portal OTICS.

O Projeto OTICS tem como objetivo geral o desenvolvimento e a implantação de soluções tecnológicas e operacionais para suporte à gestão e ao ensino. Entre essas, a criação de um ambiente na internet para a disponibilização de informações e indicadores de saúde e de ferramentas para uso educacional.

O Portal do Observatório constituiu-se em uma ferramenta de fortalecimento do uso da informação e suas tecnologias, além de disponibilizar recursos tecnológicos para uma utilização mais ampliada. Tem uso como ferramenta de cooperação técnica mais horizontal entre instituições e sistemas regionais de saúde. Este é baseado no gerenciador de conteúdo Plone ([www.plone.org](http://www.plone.org)), que disponibiliza a infra-estrutura necessária para o desenvolvimento de ambientes desta natureza. Após a criação da identidade visual para caracterizar o portal foram instaladas funcionalidades adicionais,

que permitem o fácil acesso para a utilização de fóruns de discussão, salas de bate papo, entre outros recursos.

O objetivo do projeto é disponibilizar aos parceiros e colaboradores todas as ferramentas desenvolvidas no escopo do projeto OTICS na Universidade de Caxias do Sul, servindo como suporte para o trabalho colaborativo da equipe, bem como reunir informações e análises sobre o uso de TICs na área da Saúde. Além disso, é objetivo disponibilizar os seguintes serviços no portal: videoconferência, consulta a bases de dados da área da saúde fazendo o seu cruzamento com dados demográficos (IBGE), edição e publicação de vídeos, mineração de texto, assistente virtual e Ambiente Virtual de Aprendizagem (AVA).

A 'Figura 2: Portal OTICS' mostra como é o Portal OTICS atualmente:



Figura 2: Portal OTICS

## 6. Modelo Proposto

As bases de dados e relatórios do Portal OTICS podem ser muito extensas e conter informações sobre as mais variadas áreas da saúde, tornando mais complicada a exploração de documentos e busca de informações por parte dos usuários. Se tais documentos, ao serem cadastrados, forem enquadrados dentro de determinadas áreas médicas, por exemplo, isso permitiria que o usuário navegasse por categorias semânticas de documentos, e não apenas por data de inserção e acompanhamento da UBS.

O trabalho desta pesquisa tem como objetivo o enriquecimento semântico do processo de mineração textual implementado na ferramenta Sobek, através de uma ontologia, para permitir a categorização de conteúdos dentro do Portal OTICS. O processo se dará em três estágios:

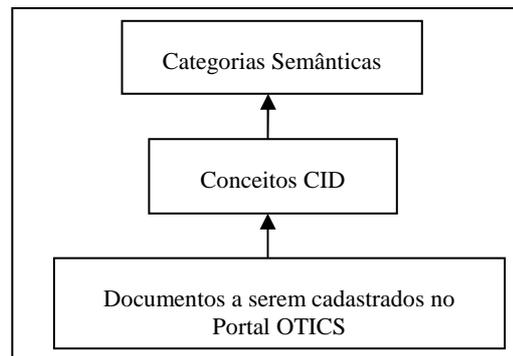
- Extração de conceitos principais dos documentos com a ferramenta Sobek;
- Comparação com CID – Classificação Internacional de Doenças;
- Enquadramento com categorias semânticas provenientes da generalização de aspectos médicos e do próprio CID.

No primeiro estágio os conteúdos a serem cadastrados no Portal OTICS serão avaliados pela ferramenta Sobek, que será responsável por extrair os termos mais relevantes encontrados no documento. Este estágio será realizado com todo material disponibilizado no portal.

No segundo estágio, a estrutura com os tópicos extraídos sobre cada um dos materiais em questão será comparada com a tabela CID, que descreve de forma padronizada todos os termos utilizados para descrição das doenças a nível internacional. Assim, cada documento será enquadrado inicialmente como pertencente a uma ou mais classificações de doenças internacionalmente conhecidas. Trata-se de um filtro inicial, em que tal vocabulário pode ser trabalhado com demais vocabulários específicos de determinadas áreas, como por exemplo, a pediatria ou geriatria.

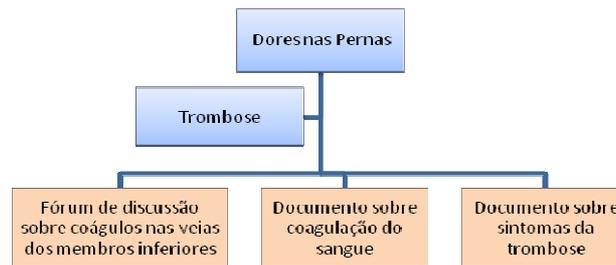
Até o final do segundo estágio é construída uma grande relação de termos específicos da Medicina com uma série de documentos relacionados aos mesmos, constituindo uma relação de N-N, em que um conceito é relacionado com N documentos a serem disponibilizados no Portal OTICS, e um documento pode ser relacionado com N conceitos. No terceiro estágio acontece o enquadramento de cada um destes conceitos CID (ou provenientes de demais vocabulários) com categorias semânticas dentro da Medicina, constituindo uma ontologia. Assim, podem existir diferentes níveis de navegação pelos conteúdos do observatório, possibilitando aos alunos, professores e profissionais usuários do portal, que explorem os diversos documentos disponíveis até encontrar o material desejado.

Em suma, são três camadas que definem o processo de interação do usuário, conforme abordado a seguir na Figura 3: Diagrama do Processo de Enquadramento Semântico:



**Figura 3: Diagrama do Processo de Enquadramento Semântico**

Abordando um cenário hipotético em que um aluno esteja explorando o Portal OTICS, pode-se considerar que este esteja consultando materiais disponíveis no observatório sobre determinada categoria semântica até que encontra um conceito que lhe interesse, 'dores nas pernas' por exemplo. Neste caso, como ilustrado na Figura 4, o portal poderia mostrar os conceitos relacionados, tal como 'trombose', presentes na classificação CID. Ao explorar este conceito, trechos de um fórum de discussão sobre o tema 'coágulos nas veias das pernas', relacionados ao tópico CID de 'trombose', poderiam ser apresentados ao aluno. Além destes, o documento sobre 'coagulação do sangue' e outro sobre os sintomas de tal enfermidade também poderiam ser relacionados.



**Figura 4: Diagrama sobre interface hipotética de exploração de conteúdos**

Através deste processo, o aluno poderia localizar facilmente materiais relevantes que seriam difíceis de encontrar através de uma simples busca. Ao navegar pelos conteúdos em diferentes níveis, agrupados em camadas distintas, o aluno poderia gradativamente estreitar sua busca até visualizar as classificações CID correspondentes aos documentos selecionados.

## 7. Considerações Finais

A exploração de materiais em ambientes virtuais tem uma lacuna a ser preenchida, que vem a ser a sua exploração sob uma perspectiva semântica, resultando em uma interação diferenciada para o usuário. A abordagem proposta oferece uma alternativa a isto, com a utilização do enquadramento semântico de extensos materiais que uma ontologia permite realizar. Além disto, a ferramenta em desenvolvimento já integra componentes desenvolvidos dentro do Projeto Sobek, otimizando a sua integração e possibilitando que testes iniciais sejam logo realizados trazendo dados para amadurecimento do projeto.

O trabalho em questão possui várias oportunidades de pesquisa posterior. A mais importante delas vem a ser a definição da ontologia que enquadrará semanticamente os termos em questão, sejam eles de um vocabulário como o CID ou outros mais. Além disto, deve-se avaliar se apenas uma ontologia será suficiente, ou se mais ontologias especializadas trabalhando em conjunto trariam um melhor resultado na exploração de conteúdos pelo usuário. Além disto, o estudo da mineração textual de tais documentos considerando classes gramaticais específicas, como verbos e substantivos, pode tornar mais eficiente o primeiro estágio comentando na seção de 'Modelo Proposto'.

## Referências

- AZEVEDO, B.; REATEGUI, E. B.; BEHAR, P. Qualitative Analysis of Discussion Forums. **IADIS International Conference on e-Learning**, Freiburg, 2010.
- BARBOSA, M. L.; REATEGUI, E. B.; SEVERO, C. E. P.; REATEGUI, E. B. Mineração de padrões no gênero textual blog. **RENOTE. Revista Novas Tecnologias na Educação**, 2009.
- BLOOM, B. S. The 2 Sigma Problem: The search for methods of group instruction as effective as one-to-one tutoring". In: \_\_\_\_\_ **Educational Researcher**. [S.l.]: [s.n.], 1984. p. 4-16.
- DALMOLIN, L.; NASSAR, S. M.; BASTOS, R. C.; MATEUS, G. P; A Concept Map Extractor Tool for Teaching and Learning. **Ninth IEEE International Conference on Advanced Learning Technologies**, 2009.



- DEVEDZIC, V. **Semantic Web and Education**. [S.l.]: Springer, 2006.
- EYHARABIDE, V. Ambientes personalizados de e-learning: considerando os contextos dos alunos. **Informática na educação: Teoria e Prática**, 2009.
- FERREIRA, F. Sumarização de Texto em Ambientes Educacionais na Web. **Simpósio Brasileiro de Informática na Educação**, 2009.
- GRUBER, T. R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. **International Journal of Human Computer Studies**, 1995.
- GUARINO, N. Formal Ontology, Conceptual Analysis and Knowledge Representation. **International Journal of Human Computer Studies**, 1995.
- ISOTANI, S.; BITTENCOURT, I. Estado da Arte em Web Semântica e Web 2.0: Potencialidades e Tendências da Nova Geração de Ambientes de Ensino na Internet. **Revista Brasileira de Informática na Educação**, v. 17, 2009.
- JACOB, E. Ontologies and the Semantic Web. **Bulletin of the American Society for Information Science and Technology**, p. 3, 2003.
- JOVANOVIC, J. et al. Leveraging the Social Semantic Web in Intelligent Tutoring Systems. **Proceedings of the International Conference on Intelligent Tutoring Systems**, 2008. 563-572.
- KLEMMANN, M.; LORENZATTI, A.; REATEGUI, E. O Emprego da Ferramenta de Mineração de Textos SOBEK como Apoio à Produção Textual. **Simpósio Brasileiro de Informática na Educação**, 2009.
- LEE, T.-B.; HENDLER, J.; LASSILA, O. The Semantic Web. **Scientific American**, 2001.
- MACEDO, A. et al. Using text-mining to support the evaluation of texts produced collaboratively. **9th World Conference on Computers in Education**, Bento Gonçalves, 2009.
- MIZOGUCHI, R.; HAYASHI, Y.; BOURDEAU, J. Inside Theory-Aware Authoring System. **Proceedings of the Int. Workshop on Ontologies and Semantic Web for E-Learning (SWEL)**, 2007.
- MURRAY, T. Authoring intelligent tutoring systems: An analysis of the state of the art. **International journal of artificial intelligence in education**, vol. 10, 1999.
- NOVAK, J.; GOWIN, D. **NOVAK, J.D., GOWIN, D.B. Learning how to learn**. New York: Cambridge University Press, 1984.
- PINHO, I. C. ; FINCO, M. D.; RIBEIRO, A. M ; REATEGUI, E. A Pedagogical Agent which Incorporates a Text Mining Tool to Promote Collaborative Writing. Em **Anais do International Conference on Frontiers in Education: Computer Science and Computer Engineering - FECS 2011**, Las Vegas, Estados Unidos. 2011.
- SCHENKER, A. **Graph-Theoretic Techniques for Web Content Mining. PhD thesis, 2003**. University of South Florida. [S.l.]. 2003.
- SHARP, M. **Text Mining**. [S.l.]: Rutgers University, 2001.
- SOY, S. Automatic Indexing References to the Early Years of Automatic Indexing and Information Retrieval, 2003.  
Disponível em: <<http://www.gslis.utexas.edu/~ssoy/organizing/l391d2c.htm>>.
- WITTEN, I. H.; FRANK, E. **Data Mining: Practical Machine Learning Tools and Techniques**, 2005. San Francisco: Morgan Kaufman, 2005.
- ZHOU, L. Ontology Learning: state of the art and open issues. In: **SPRINGER Information Technology and Management**. [S.l.]: [s.n.], 2007.