

Mineração de dados educacionais para a construção de alertas em ambientes virtuais de aprendizagem como apoio à prática docente

Adriana Justin Cerveira Kampff - akampff@gmail.com^{1,2}

Eliseo Berni Reategui - eliseoreategui@gmail.com¹

José Valdeni de Lima - valdeni@inf.ufrgs.br¹

¹PGIE – UFRGS

²ULBRA – Tecnologia e Computação / EAD

Resumo. *O Ensino a Distância no país apresenta-se em franca expansão. Cresce o número de alunos e gerenciar seus processos de aprendizagem em Ambientes Virtuais de Aprendizagem (AVA), com qualidade de interações e de acompanhamento, exige cada vez mais do professor. Este artigo relata o estágio atual de uma pesquisa que busca, através de mineração de dados, identificar perfis de alunos com risco de evasão ou reprovação, bem como boas práticas pedagógicas de acompanhamento e intervenção, de forma a promover alterações nos AVA que facilitem a implementação de alertas. Tais alertas serão gerados para o professor a partir de agrupamentos de alunos com necessidades similares, para que ele possa dirigir uma comunicação ou atividade pessoal e contextualizada para esses sujeitos, tornando sua mediação mais efetiva.*

Palavras-chaves: *Ambientes Virtuais de Aprendizagem, Mineração de Dados, Sistemas de Alertas.*

Educational Data Mining to support Teacher in Learning Management Systems

Abstract. *Distance learning is expanding in Brazil. There is a very large number of students nowadays attending such courses, and to manage their learning processes in a Virtual Learning Environments requires a lot of effort from the teachers. This article reports the current stage of a research project which intends to use data mining techniques to identify students' profiles that have the risk of evading or getting bad grades. The goal is to promote best teaching practices for monitoring and interacting with students. The alerts will be generated to the teacher by suggesting groups of students with similar needs, so that the teacher can do a personal and contextualized communication (or activity) for these students, making mediation more effective.*

Keywords: *Learning Management Systems, Data Mining, Awareness Systems.*

1. Introdução

A Educação a Distância (EAD) apresenta-se hoje como a modalidade que permite levar a educação a locais distantes, flexibilizar os horários de estudos e envolver um número maior de sujeitos. É uma realidade aceita e estimulada por órgãos governamentais (Costa 2007), também amplamente explorada pela iniciativa privada. Mas, não basta democratizar o acesso: é preciso garantir a qualidade da educação ofertada. Bons professores, materiais fartos e interativos, alunos autônomos e motivados são elementos importantes para o sucesso nessa modalidade de ensino (Peters, 2001).

Cresce, a cada dia, a oferta de cursos a distância. No Ensino Superior, cursos de graduação são ofertados de forma integral na modalidade EAD. Também em cursos de graduação presenciais 20% de suas disciplinas podem ser realizadas a distância, conforme legislação vigente (MEC/Portaria n. 4.059, de 10 de dezembro de 2004.). Embora a idéia de que cursar uma graduação, total ou parcialmente, a distância, traz flexibilidade de tempo e local de estudo para o aluno, diversas pesquisas mostram a dificuldade do aluno em manter uma rotina de estudos que possibilite seus avanços, elevando os índices de evasão (Favero & Franco 2006).

Frente às novas necessidades educacionais, professores e alunos precisam reaprender a exercer seus papéis. Os professores devem abandonar de uma vez por todas a idéia de transmissão de conhecimento, já que os conteúdos são postos para o aluno em um momento definido, e devem passar a planejar suas “aulas”, ou melhor, suas “interações”, como mediação dos alunos na construção do conhecimento (Passerino et al, 2007). Os alunos, por sua vez, precisam desenvolver novas atitudes de busca e acompanhamento das aulas, com uma postura mais autônoma e participativa. Professores e alunos não habituados a essa modalidade de ensino podem valer-se de ferramentas, agentes ou assistentes virtuais que os auxiliem nas tarefas educacionais (Kampff, 1999; Reategui et al, 2008).

O professor que acompanha uma turma de alunos em EAD deve transcender o papel de conteudista, ressaltando que, muitas vezes, o material didático já foi elaborado por outro professor. O que fará, então, esse professor? Ficarà a disposição apenas para esclarecer dúvidas surgidas nos auto-estudos dos alunos? Definitivamente, apenas reagir à demanda dos alunos, apresentando respostas às suas indagações, empobrece o papel do professor. O professor, com postura pró-ativa e medidora, deverá gerir a sala de aula virtual de forma a integrar seus alunos em uma comunidade virtual de aprendizagem, que estabeleça trocas significativas e onde cada um esteja engajado nos estudos.

O professor mediador deverá estabelecer práticas que permitam aproximar os alunos (entre si e dos objetos de estudo), embora geograficamente distantes, engajando-os em seus processos de aprendizagem. É preciso acompanhar de perto os processos de aprendizagem, oferecendo caminhos para aprendizagens mais efetivas aos alunos, ajudando-os a estabelecer rotinas de estudos e posturas de buscas, oportunizando avaliações formativas com *feedback* e estudos adicionais.

O foco no trabalho aqui descrito é o professor: auxiliá-lo na condução de sua ação mediadora por meio de mecanismos de conscientização (alertas) sobre situações problemáticas que se apresentam nos ambientes virtuais de aprendizagem, nem sempre facilmente percebidas no cotidiano da ação pedagógica. O contexto das interações abordadas ocorre em disciplinas totalmente a distância em cursos de graduação, por meio de ambientes virtuais de aprendizagem (AVA).

2. Metodologia

Para conceber um conjunto de alertas (notificações automáticas) que auxilie o professor na sua ação mediadora, é preciso compreender como os sujeitos aprendem e constroem conhecimento. Essa compreensão permite embasar a elaboração dos materiais e dos possíveis percursos didáticos, e que ações o professor pode realizar na mediação da aprendizagem, a partir das ferramentas disponíveis nos ambientes virtuais utilizados para atividades de EAD. Técnicas de mineração de dados podem auxiliar na localização de perfis de alunos (a partir da análise de dados gerados no AVA, em seus diversos espaços de informação e comunicação) que possibilitem mapear comportamentos comuns que levem à evasão ou reprovação.

A partir disso, torna-se possível propor um conjunto de alertas a serem implementados que efetivamente promovam ganhos pedagógicos, tornando o professor mais eficiente no acompanhamento e mediação da turma e, conseqüentemente, que essa ação docente repercuta em alunos mais engajado na sua aprendizagem, que se sintam apoiados e seguros em seus percursos. Para desenvolver o trabalho proposto, diversos passos vêm sendo desenvolvidos e pesquisados:

- O processo de capacitação docente para a EAD: preparar e atuar na formação de professores que, posteriormente, atuarão nessa modalidade requer que se reconheçam os diversos papéis docentes atualmente existentes nessa modalidade (conteudista / especialista / tutor), vivenciando a produção do material didático (escrito ou audiovisual), a construção de planos de aula inovadores e consistentes, as ferramentas disponíveis e a forma de mediação em AVA, motivando e acompanhando os alunos;
- O acompanhamento da prática docente em EAD: por meio de suporte à montagem, de análise e de discussão dos planos de aula, bem como reflexões a partir da atuação do professor no AVA;
- A realização de entrevistas com professores, buscando as melhores práticas pedagógicas: coletando dados sobre formas adequadas de mediação, a partir da percepção e experiência dos professores entrevistados, com a finalidade de apoiar o desenvolvimento dos alertas;
- A mineração de dados educacionais disponíveis nos AVA, analisando quantidade e frequência de acessos e ações aos recursos de informação e comunicação disponíveis e suas relações com desempenho: evasão, reprovação e aprovação.
- A definição dos alertas: objetivando apoiar o acompanhamento dos professores experientes, além de apresentar aos professores iniciantes as melhores práticas já mapeadas e pontos de atenção a partir dos dados minerados;
- A implementação e validação dos alertas: para tanto, a equipe de desenvolvimento do AVA utilizado pelo grupo de professores pesquisado já está realizando as alterações nas ferramentas, de maneira a permitir um monitoramento mais efetivo e a geração desses alertas.

Quando se utiliza um AVA, não basta disponibilizar conteúdos e abrir espaços de discussão, é preciso que o professor tenha uma postura de mediação, convidando os alunos a participarem e acompanhar seus desempenhos, dando retornos e orientações durante toda a caminhada. O professor deve empregar estratégias que contribuam para o envolvimento dos alunos nos estudos propostos, propiciando que busquem também

informações complementares e contribuam com o grupo, sentindo-se pertencentes à comunidade de aprendizagem, reconhecidos e valorizados.

3. Mineração de Dados Educacionais

A partir da metodologia proposta, o estágio atual do trabalho está em identificar e propor um conjunto de alertas para ambientes virtuais de aprendizagem que auxiliem o professor na gestão das situações de aprendizagem. Para um acompanhamento mais efetivo dos alunos de uma turma em EAD e buscando minimizar a sobrecarga de trabalho dos professores, torna-se desejável valer-se de dados históricos de interação em AVAs, buscando descobrir conhecimento potencialmente útil, de forma a agregar mais informações para a geração de alertas automatizados.

Imensos volumes de dados são gerados pela interação de usuários (professores e alunos) em AVAs, sendo sistematicamente armazenados e, muitas vezes, recuperados. Mas, frequentemente, dados de logs, detalhados e extensos, são difíceis de interpretar e acabam sendo deixados de lado no cotidiano pedagógico. É preciso analisar esses dados de forma a descobrir conhecimento que possa auxiliar na resolução de problemas educacionais. Técnicas de Mineração de Dados (*Data Mining*) podem ser empregadas para descobrir comportamentos ou características de alunos com maior propensão à evasão ou reprovação e, então, essa descoberta pode contribuir para a tomada de decisão de como atuar de forma a evitar esses resultados indesejados.

Além das tabelas com os dados gerados nas diversas ferramentas (datas e quantidades de acessos aos conteúdos e arquivos nas bibliotecas virtuais; acessos às atividades de avaliações, datas de entrega, acertos e notas; envio de e-mails, com identificação de datas e remetentes; participações em fóruns, com acessos de leitura e postagens...), pretende-se avaliar as relações dessas informações com outras como o sexo, a idade, a unidade presencial (pólo) e o curso de origem do aluno, analisando disciplinas básicas ofertadas totalmente a distância para alunos de cursos variados de graduação. A partir desses dados, também são consideradas as notas desses alunos para verificar a existência de relações entre os comportamentos e os desempenhos.

Esses dados devem ser tratados para que possam ser submetidos às ferramentas que implementam técnicas de Mineração de Dados. As etapas para Descoberta de Conhecimento consistem em: selecionar os dados, pré-processá-los para retirar dados duplicados ou inconsistentes, transformá-los nos formatos adequados conforme algoritmos e ferramentas utilizadas, submeter os dados transformados à mineração propriamente dita, e, após, fazer a interpretação dos resultados (Fayyad et al, 1996). Cabe destacar que esse processo, sempre que necessário, retorna a fases anteriores, visando a melhorar os resultados.

Para realizar a mineração de dados, então, selecionam-se os dados relevantes e preparam-se esses dados para a mineração, excluindo dados redundantes e uniformizando a forma de representação dos mesmos que, originalmente, podem estar armazenados em tabelas diferentes, com formas de representação também diferenciadas. Após a preparação dos dados, é preciso definir as ferramentas e os algoritmos de mineração a serem utilizados.

Lopes & Schiel (2008) modelaram um sistema para apoio à decisão (SAD) para auxiliar professores a identificar perfis de estudantes em EAD. Na etapa de seleção de dados, definiram um conjunto de análise denominado MAD – Modelo Analítico de Dados, que engloba número de e-mails enviados, quantidade de acesso aos recursos didáticos, nota em cada item de avaliação, estado (UF), sexo, idade e situação

(cursando, trancado ou desistente). Para a mineração, foi utilizando o *framework* WEKA, desenvolvido por Witten & Frank (2000 apud Lopes & Schiel, 2008). O sistema definido minerou dados fictícios, sem uma validação com dados reais.

No nosso trabalho, um experimento realizado tomou por base a mineração de dados reais de 161 estudantes, de duas turmas de uma mesma disciplina, ofertada na modalidade EAD, para alunos de cursos de graduação presencial. A disciplina é semestral e os dados coletados referem-se aos três primeiros meses de estudo, até o fechamento da primeira nota utilizada para compor a nota final na disciplina - a nota final dos alunos é uma média ponderada entre a primeira nota com peso um e a segunda nota com peso dois.

Os alunos são de três pólos diferentes (unidades presenciais onde ocorrem a aula introdutória e as provas finais, na proporção de 139/11/11) e de cinco áreas do conhecimento diferentes (conforme agrupamento da Universidade pesquisada, 55/51/29/20/5). Entre os alunos da amostra, 78 são do sexo masculino e 83 do sexo feminino; 92 alunos têm idade entre 16 e 25 anos, 52 entre 26 e 35 anos, 11 entre 36 e 45 anos e 6 possuem 46 anos ou mais. Considerou-se a época do primeiro acesso ao AVA (primeira, segunda ou terceira quinzena da disciplina), a época do último acesso (semana corrente, semana passada, há mais de quinze dias ou há mais de um mês), o número de acessos ao AVA (que variou de zero a 80), o número de acessos à biblioteca virtual (que variou de zero a 50), o número de postagens em fóruns (de zero a 12) e o número de avaliações entregues (de zero a três). Buscou-se relacionar esses dados com as notas obtidas pelos alunos durante os três meses de acompanhamento da disciplina. Os dados preparados foram convertidos para o formato CSV (*Comma Separated Values*) e submetidos à ferramenta *RapidMiner*¹. A figura 2 apresenta a visualização dos metadados.

Type	Name	Value Type	Statistics	Range	Unk...
label	Faixa	nominal	mode = Nota [7..9] (79)	Nota [0..5] (41), Nota [9..10] (12), Nota [7..9] (79), Nota [5..7] (29)	0
regular	Sexo	nominal	mode = F (83)	M (78), F (83)	0
regular	Pólo	nominal	mode = Canoas (139)	Canoas (139), SantaMaria (11), Gravataí (11)	0
regular	Área	nominal	mode = Educação,CiênciasEArtes (55)	SaúdeEBemEstarSocial (51), TecnologiaEComputação (20), Educação,CiênciasEArtes (55), CiênciasSociaisAplicadas (29), CiênciasAgrárias (6)	0
regular	FaixaEtária	nominal	mode = Idade: [16..25] (92)	Idade: [36..45] (11), Idade: [26..35] (52), Idade: [16..25] (92), Idade > 45 (6)	0
regular	PrnAc1	nominal	mode = Quinzena1 (113)	Quinzena1 (113), Quinzena4 (1), NC (4), Quinzena2 (32), Quinzena3 (10)	1
regular	ÚtAc1	nominal	mode = EssaSemana (114)	MaisDeUmMês (5), EssaSemana (114), NC (4), De7a15DiasAtrás (23), De15a30DiasAtrás (15)	0
regular	NroAces	integer	avg = 20.989 +/- 15.140	[0.000 ; 83.000]	0
regular	T.Fór	integer	avg = 2.950 +/- 2.793	[0.000 ; 12.000]	0
regular	T.Bib	integer	avg = 12.224 +/- 8.946	[0.000 ; 50.000]	0
regular	T.Ava	nominal	mode = 3 (121)	0 (12), 3 (121), 1 (5), 2 (23)	0

Figura 2 – RapidMiner: visualização dos metadados

A partir dos dados tratados, foram utilizados algoritmos disponíveis na ferramenta *RapidMiner* para aprendizagem supervisionada, gerando regras no formato Se <condição> Então <resultado>. O foco era descobrir a relação entre comportamentos dos alunos – como acesso ao ambiente e às leituras sugeridas, entrega das atividades, postagens em fóruns – e características demográficas – como faixa etária, sexo, curso e pólo de origem – e suas possíveis repercussões nos desempenhos dos alunos. Aplicou-se, também, um algoritmo para a geração de Árvores de Decisão, alterando-se os dados de desempenho (ao invés de faixas de valores, trabalhando com os conceitos de acima

¹ <http://rapid-i.com/content/blogcategory/38/69/> Acessado em: 28/10/2008.

ou abaixo na média) que apresenta os resultados de maneira gráfica e textual, como apresentado nas figuras 3 e 4.

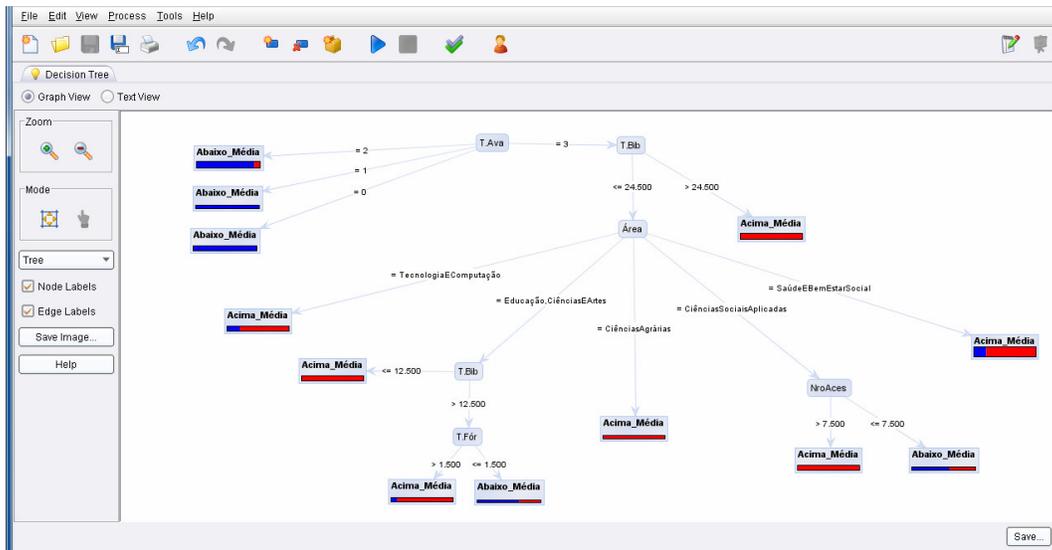


Figura 3 – Visualização gráfica na Árvore de Decisão gerada

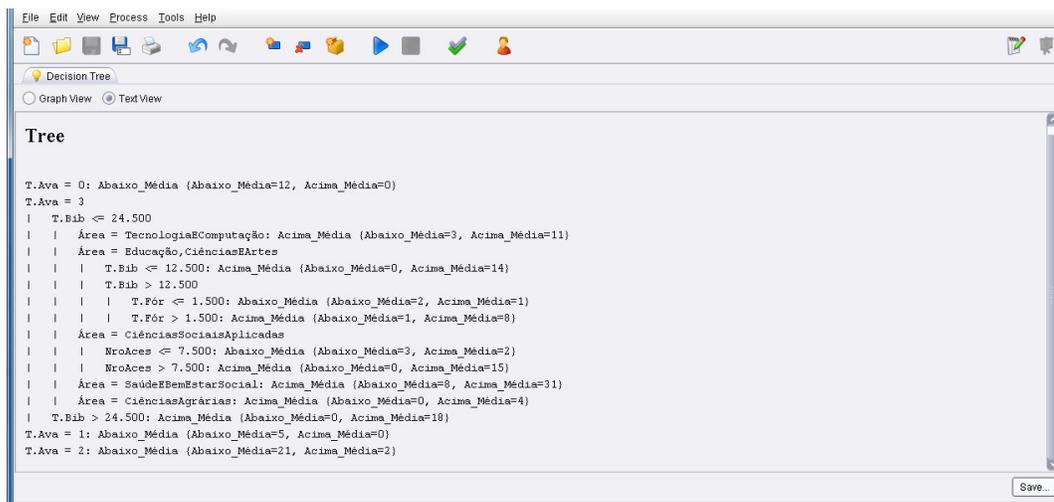


Figura 4 – Visualização textual na Árvore de Decisão gerada

Ao interpretar a Árvore de Decisão gerada, observa-se que dificilmente os alunos que não realizaram todas as atividades avaliativas propostas obtiveram nota acima da média – reforça-se a importância de engajar e envolver os alunos na realização das atividades. A análise mais relevante refere-se, portanto, aos alunos que realizaram todas as tarefas propostas (T.Ava=3). Desses, todos os alunos com mais de 24 acessos aos materiais de estudo, periodicamente postados na Biblioteca Virtual do AVA utilizado, obtiveram bons resultados (Acima_Média). Entre os alunos da Área de Educação, Ciências e Artes, por exemplo, é possível observar que, para aqueles que tiveram entre 12 e 25 acessos à Biblioteca Virtual, o que distinguiu os alunos acima ou abaixo da média foi o volume de postagens realizadas nos fóruns de discussão: os alunos acima da média tiveram, no mínimo, duas postagens nos fóruns.

Observou-se, a partir da mineração, o potencial de analisar os dados gerados em AVAs para descoberta de conhecimento. Para resultados mais efetivos, pretende-se ampliar a amostra. Serão acrescentados ainda dados que permitam acompanhar a periodicidade de acesso às diversas ferramentas existentes no AVA utilizado pelos alunos, ampliando a seleção de dados a outras ferramentas de comunicação não consideradas nesse projeto inicial, além de uma análise mais detalhada das atividades avaliativas, como por exemplo questões de maior impacto no desempenho final de cada avaliação. É preciso dispor de um volume de dados maior do que o inicialmente utilizado para a efetiva localização de padrões que possam subsidiar a atuação do professor.

Quanto mais informações estiverem à disposição para análise, potencialmente mais rico será o resultado da mineração. Em relação a materiais disponibilizados, por exemplo, pelo professor ou por outros alunos, se o AVA aceita comentários a respeito de cada um dos materiais, bem como conta com uma categoria de qualificação, essa valoração pode vir a ser ponderada em relação aos acessos a tais materiais, ou seja, o acesso a materiais bem-avaliados pode ser mais importante do que outros cuja avaliação não é tão positiva. Analisando individualmente um aluno, o professor pode avaliar não apenas a quantidade de materiais compartilhados, mas a qualidade desses a partir da avaliação participativa dos colegas de turma. Quando debate-se por meio de ferramentas de fóruns de discussão, além da quantidade de postagens, é fundamental analisar o conteúdo dessas postagens, atribuindo valor a cada uma, que possibilite filtrá-las por importância ou qualidade da reflexão.

O objetivo final da mineração de dados nesse projeto é servir de base para a construção de alertas que conscientizem o professor sobre padrões de comportamentos de seus alunos, chamando a atenção do professor sobre situações que podem levar à evasão ou reprovação.

4. Considerações Finais

Além da pesquisa de boas práticas educacionais e de entrevista junto a professores de EAD sobre o acompanhamento que julgam ideal oferecer aos alunos, a partir de suas vivências e concepções epistemológicas, o sistema de alertas em desenvolvimento será alimentado por mineração de dados, que auxiliam na descoberta de conhecimento potencialmente útil e não facilmente percebido nos dados da forma como se encontram armazenados nos AVAs. Alertas automatizados, embasados em boas práticas pedagógicas e mineração de dados, podem contribuir para um acompanhamento mais efetivo dos alunos, em um panorama de cursos cada vez mais difundidos e realidades cada vez mais heterogêneas.

Referências Bibliográficas

- COSTA, José da. Modelos de Educação Superior a Distância e Implementação da Universidade Aberta do Brasil. **Revista Brasileira de Informática na Educação** – V. 15 N. 2 - Maio a Agosto, 2007.
- FAVERO, Rute Vera Maria; FRANCO, Sérgio Roberto Kieling. Um estudo sobre a permanência e a evasão na Educação a Distância. **Revista Novas Tecnologias na Educação** – V. 4 N.º 2 – Dezembro, 2006.
- FAYYAD, U. M.; PIATESKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery: An Overview. **Advances in Knowledge Discovery and Data Mining**, AAAI Press, 1996.



- KAMPPFF, Adriana Justin Cerveira Kampff. **Utilização de Agentes para Suporte ao Trabalho em Grupo na Internet**. Porto Alegre: PGCC/UFRGS. 1999. Dissertação de Mestrado.
- LOPES, Claudivan Cruz; SCHIEL, Ulrich. Uma Estratégia para Aplicar Mineração de Dados no Acompanhamento do Aprendizado na EaD. Disponível em: <http://www.inf.furb.br/seminco/2004/artigos/108-vf.pdf>. Acessado em: 20/11/2008.
- MEC/Portaria. **Nº 4.059, de 10 de dezembro de 2004**. Disponível em: http://portal.mec.gov.br/sesu/arquivos/pdf/nova/acs_portaria4059.pdf. Acessado em: 24/08/2008.
- PASSERINO, Liliana M.; GLUZ, João Carlos; VICARI, Rosa Maria. **MEDIATEC – Mediação Tecnológica em Espaços Virtuais para Apoio ao Professor Online**. **Revista Novas Tecnologias na Educação**, v. 5, p. 1, 2007.
- PETERS, Otto. **Didática do Ensino a Distância**. São Leopoldo: Unisinos. 2001.
- REATEGUI, Eliseo Berni et al. Um Sistema Multiagente para Controle de um Assistente Pessoal Aplicado a um Ambiente Virtual de Aprendizagem. **Revista Novas Tecnologias na Educação – V. 6 N.º 1 – Julho, 2008**.