

## MINERAÇÃO DE DADOS PARA APOIO À TOMADA DE DECISÃO DE DOCENTES NA AVALIAÇÃO DE ALUNOS DE GRADUAÇÃO A DISTÂNCIA

*José Leonardo Oliveira Lima*

**Resumo:** A presente comunicação é uma síntese do projeto de Doutorado do curso de Ciência da Informação e relato da fase inicial em que se encontra a pesquisa. Objetiva propor um processo de mineração de dados e desenvolver um módulo (protótipo) de apoio à tomada de decisão para avaliação, a ser integrado a um Ambiente Virtual de Aprendizagem (AVA), satisfazendo as necessidades de informação dos docentes na avaliação de alunos em cursos de graduação a distância. Os procedimentos metodológicos envolvem pesquisa descritiva, metodológica e aplicada com pesquisa bibliográfica e de campo. A pesquisa se encontra na fase inicial de revisão bibliográfica, construção do referencial teórico e comunicação com os pares e especialistas nos assuntos objetos de estudo, mas já com elementos relevantes sobre a temática.

**Palavras-chave:** Necessidades e uso da informação, mineração de dados, avaliação, educação a distância.

### 1 INTRODUÇÃO

A humanidade tem vivenciado um período histórico de mudanças e incertezas. Com o volume de informações e recursos disponíveis, o ser humano e as organizações apresentam sinais de dificuldade, confusão, frustração e impotência quanto ao uso do volume de dados e recursos tecnológico disponíveis, evidenciando a distância ainda existente entre a disponibilização de novas tecnologias ou recursos e a capacidade para o uso que possibilitem o acesso e a recuperação da informação com eficiência e eficácia.

A universidade, como o lócus do desenvolvimento e disseminação do conhecimento científico, é um espaço privilegiado para problematizar e também absorve os benefícios e impactos das referidas mudanças.

No Brasil, na última década, ocorreu um expressivo crescimento da educação superior na modalidade a distância, mesmo com um atraso de mais de 30 anos em relação a outros países (LITTO; FORMIGA, 2009; SANCHEZ, 2008). Na Educação a Distância (EaD) “em linha” - que utiliza os recursos e possibilidades das Tecnologias da Informação e Comunicação (TICs) - os Ambientes Virtuais de Aprendizagem (AVA) favorecem a interação entre alunos e professores e possibilitam a colaboração e discussão de idéias, disponibilização de tarefas, atividades e materiais de estudo (textos, fotos, vídeos, animações etc), avaliações, construção colaborativa por meio de fóruns, áudio e videoconferência etc (MOORE; KEARSLEY, 2007).

As TICs podem ser elementos potencializadores da atividade docente e importante apoio ao processo ensino-aprendizagem, mas também oferecem diversos desafios. Por exemplo: como avaliar efetivamente a participação do aluno, a originalidade dos materiais que o aluno posta, a qualidade e efetividade da sua participação em um AVA?

Do mesmo modo que existe a Internet para favorecer o acesso à informação, também existem os mecanismos usados pelos alunos para cometer pequenos delitos, como o “Control C + Control V”, os sites de disponibilização de trabalhos prontos etc. Esses são apenas alguns exemplos que permitem ilustrar a problemática que envolve a atividade docente de avaliação<sup>1</sup> dos alunos que utilizam os AVAs para participação, interação e resolução de atividades nos cursos a distância.

Muitos dos AVAs apresentam relatórios com informações sobre cada aluno, acesso aos recursos do sistema, *logs*<sup>2</sup> de uso do ambiente etc, porém essas são muito detalhadas, técnicas e confusas. Faltam informações sintéticas e significativas (inclusive com gráficos e elementos visuais) que facilitem ao professor detectar não conformidades e poder avaliar, de forma mais justa e consistente, a participação e o desenvolvimento do conhecimento do aluno em um curso por intermédio de um AVA.

Face ao exposto, temos como problema de pesquisa: quais são as necessidades de informação dos docentes dos cursos de graduação a distância no processo de avaliação dos alunos em um Ambiente Virtual de Aprendizagem (AVA) e como a mineração de dados (*data mining*) pode ser integrada para apoiar à tomada de decisão nesse processo de avaliação?

O objetivo geral do presente projeto é propor um processo de mineração de dados e desenvolver um módulo (protótipo) de apoio à tomada de decisão para avaliação, a ser integrado a um AVA, satisfazendo as necessidades de informação dos docentes na avaliação de alunos em cursos de graduação a distância.

## 2 REFERENCIAL TEÓRICO

Capurro e Hjørland (2007) colocam que a “sociedade da informação” passou a ser assim denominada em razão do surgimento da Tecnologia da Informação (TI) e dos seus impactos globais. Referidos autores também apresentam a dificuldade de se manter o termo “informação” formalmente definido, devido a sua utilização em diversos outros campos do conhecimento.

Segundo Capurro e Hjørland (*op. cit.*, 154-155): “quando usamos o termo informação em Ciência da Informação (CI), deveríamos ter em mente que informação é o que é informativo para uma determinada pessoa”. Para o presente projeto, definimos o termo “informação” considerando essa perspectiva.

---

<sup>1</sup> Avaliação é um processo amplo e complexo que lida com diversos aspectos do processo ensino-aprendizagem. Para fins do presente projeto, consideramos os elementos disponíveis no AVA (dados, logs, postagens, participação etc.) que possam ser minerados, sintetizados e integrados e que sirvam de apoio ao processo de avaliação e tomada de decisão pelo professor para compor uma nota ou menção de avaliação, que poderá ser formativa ou somativa.

<sup>2</sup> Logs – arquivo de registro automático de todas as operações realizadas por um usuário em um sistema ou ambiente.

A ênfase interdisciplinar é um aspecto que apresenta a importância da colaboração entre as diversas áreas de conhecimento e é preponderante na CI. Para Le Coadic (2004, p. 20), “a ciência da informação é uma das novas interdisciplinas [que colaboram entre si] a psicologia, linguística, sociologia, informática, matemática, lógica, estatística, eletrônica, economia, direito, filosofia, política e telecomunicações”. Saracevic (1995, p. 2) apresenta a abordagem complementar entre CI e as diversas ciências, dentre elas a Ciência da Computação (CC): “Ciência da Computação aborda os algoritmos relacionados à informação, enquanto Ciência da Informação abrange a natureza da informação e sua utilização pelos seres humanos. Os dois problemas não estão em competição, eles são complementares”.

Na CI, temos os estudos que envolvem a organização da informação e a complexidade dos mecanismos de armazenamento, disponibilização e recuperação da informação, em que precisa ser observado, desde os aspectos objetivos da informação-como-coisa (que a Tecnologia da Informação dá uma maior ênfase), passando pelos aspectos da informação-como-processo e informação-como-conhecimento, conforme as nomenclaturas propostas por Buckland (1991).

Capurro e Hjørland (2007, p. 193) fazem uma distinção importante entre informação como um objeto ou coisa (por exemplo: número de bits, dados armazenado em arquivos, banco de dados, dentre outros) e informação como conceito subjetivo, dependendo da interpretação de um sujeito cognitivo.

Na CI, desde meados da década de 80, os estudos de usuários enfatizam essa mudança de enfoque, saindo do “foco no sistema” para o “foco no uso”. Dervin e Nilan (1986) levantaram as seguintes questões: sistemas de informação poderiam servir melhor os usuários – aumentando sua utilidade (do sistema) para os clientes e ser mais responsável para com eles; para servir melhor a clientela, suas necessidades e uso de informação devem ser o objetivo central do sistema; um melhor atendimento à clientela pode exigir a implementação de um sistema com orientações remodeladas; os sistemas de informação não se utilizam da tecnologia para atender melhor à clientela.

Courtright (2006) coloca que a Tecnologia da Informação (TI) está envolvida em práticas de informação do dia-a-dia, mediando e modelando não só o tipo, o volume e a apresentação da informação disponível, como também as expectativas dos atores em relação ao tipo de informação que eles podem, devem e irão procurar.

O processo de mineração de dados (*data mining*), dentre outras possibilidades da CC, pode contribuir para oferecer informações significativas para o usuário. Fayyad et. al. (1996, p. 4) definem mineração de dados como “o processo de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis, embutidos nos dados”. Consiste em um conjunto de ferramentas e técnicas que, por intermédio do uso de algoritmos, são capazes de explorar uma grande base de dados, auxiliando o homem a extrair as informações necessárias à obtenção do conhecimento (GOLDSCHMIDT; PASSOS, 2005).

A proposição de estudos de mineração de dados em um AVA envolve compreender a complexidade e a amplitude dos aspectos objetivos existentes nos dados armazenados no sistema, a



correlação entre eles para produzir novas informações que atendam as expectativas e necessidades do usuário de forma a propiciar o conhecimento (elemento subjetivo que ocorre no sujeito cognoscente).

Um Sistema de Apoio à Decisão (SAD) é definido por Stair (2005, p. 22) como um “conjunto organizado de pessoas, procedimentos, software, base de dados e dispositivos usados para dar apoio à tomada de decisões em problemas específicos”. Os elementos essenciais de um SAD incluem um conjunto de modelos usados para apoiar um tomador de decisão ou usuário (base de modelos), um conjunto de fatos e informações para apoiar à tomada de decisão (base de dados) e sistemas e procedimentos (interface com o usuário) que auxiliam os tomadores de decisão e outros usuários a interagir com o SAD.

Partindo das necessidades de informação dos docentes, das possibilidades da Mineração de Dados e da ênfase dos SADs é que propomos realizar uma síntese que denominamos “Sistema de Apoio à Decisão nos Processos de Avaliação”, a ser incorporado a um AVA, cujo processo e produto são os objetos do presente estudo.

### 3 METODOLOGIA

Utilizamos como referencial a taxionomia proposta por Vergara (2009), que apresenta o tipo da pesquisa quanto aos fins e quanto aos meios. Quanto aos fins a pesquisa é descritiva, metodológica e aplicada. Descritiva por que contempla a interpretação e descrição das percepções dos atores a luz do problema de pesquisa. Metodológica, pois envolve a proposição e desenvolvimento de um processo de mineração de dados em um AVA. Aplicada já que, ao final, será desenvolvido um módulo (protótipo de SAD) para ser integrado ao AVA. Quanto aos meios, a pesquisa é bibliográfica e de campo. Bibliográfica, para a construção do referencial teórico. De campo, pois envolve pesquisas com diversos atores dos cursos de graduação a distância que utilizam o AVA.

Os sujeitos da pesquisa (atores) são os especialistas em avaliação, professores (ou tutores), coordenadores de curso e alunos de cursos de graduação a distância. Com os especialistas, professores, tutores e coordenadores de curso, levantaremos os aspectos referentes ao processo de avaliação de alunos e as necessidades de informação para a tomada de decisão sobre o processo de avaliação, bem como o comportamento no AVA. Com os alunos serão observados os mecanismos e o comportamento na utilização do AVA para levantamento dos costumes, boas e más práticas no uso do AVA, processo de interação etc.

A amostra tem caráter não probabilístico, tendo como critério de escolha a facilidade de acesso aos participantes. São 3 cursos de graduação a distância previstos e, em cada um, contaremos com a participação de coordenadores, professores e/ou tutores (conforme as peculiaridades de cada projeto pedagógico e de cada curso) e alunos. O AVA escolhido para o presente estudo foi o Moodle pois, além de ser um software livre e de código aberto - podendo ser adaptado e personalizado conforme as necessidades de cada instituição, tem sido adotado por diversas Instituições de Educação Superior brasileiras.



Os instrumentos utilizados para a coleta dos dados são entrevistas (abertas e por pauta, dependendo da etapa e do ator), questionários (com questões objetivas e algumas abertas) e observação. Também estão sendo avaliados, para utilização, sistemas de software e ferramentas de mineração de dados (como por exemplo o WEKA), banco de dados MYSQL, linguagem de programação PHP (usada no AVA Moodle), geradores de gráficos e programa estatístico Sphinx Lexica.

Os procedimentos para a coleta de dados são os específicos para cada instrumento. O tratamento de dados está sendo feito de forma quantitativa e qualitativa. Na abordagem qualitativa, estamos utilizando os processos de análise e síntese tanto no estudo bibliográfico, em andamento, como também serão utilizados na análise dos discursos colhidos nas entrevistas e questionários com os diversos atores. Na abordagem quantitativa (ainda não iniciada), faremos o uso das ferramentas e técnicas estatísticas (frequência média, moda, desvio padrão), principalmente na redução dos questionários, bem como nas análises dos *logs* e dados referentes ao comportamento de uso de professores e alunos no AVA, e na montagem do processo de mineração de dados e módulo (protótipo) de suporte à decisão sobre avaliação.

#### 4 CONSIDERAÇÕES FINAIS

Mesmo no primeiro semestre de execução, já percebemos, pelo levantamento bibliográfico realizado e estudos da fundamentação teórica em CI, a relevância do presente estudo para além do tema delimitado, podendo ser mais um elemento contributivo no processo de desenvolvimento do conhecimento tanto nos aspectos teóricos e interdisciplinares da CI, em conjunto com a CC e Ciência da Educação, bem como nos aspectos práticos de pesquisa aplicada, contribuindo para a superação do paradoxo que foi apresentado no relatório da Conferência Mundial sobre Educação Superior realizado pela UNESCO há mais 12 anos em Paris, mas ainda existente:

A comunidade científica é a origem de grande parte das soluções técnicas e metodológicas necessárias para o desenvolvimento das Novas tecnologias da Informação e comunicação - NTIC. Contudo, os setores da educação são os que menos aproveitam das contribuições das novas tecnologias. (CONFÉRENCE MONDIALE SUR L'ENSEIGNEMENT SUPÉRIEUR, 1998, p. 7, tradução nossa).

**Abstract:** This paper is a summary of a PhD project in Information Science and a reporting of the initial researches. It contemplates the use of data mining in Virtual Learning Environment (VLE) to achieve information needs of teachers when evaluating students of undergraduate distance courses in the VLE. It proposes to develop a process and a module (prototype software built into the VLE) to perform data mining on databases in order to compose a Decision Support System (DSS) covering the relevant aspects of evaluation in a distance learning course by teachers. The methodological procedures involves descriptive, methodological and applied research, literature review and field procedures. The research is in process of literature review and theoretical construction, communication with peers and experts, but already with relevant datas.

**Keywords:** Information needs and use, data mining, evaluation, distance learning.



## REFERÊNCIAS

BUCKLAND, M. K. Information as thing. **Journal of the American Society of Information Science**, v. 42, n.5, p. 351-360, 1991.

CAPURRO, R.; HJØRLAND, B. O conceito de informação. **Perspectivas em ciência da informação**, v. 12, n. 1, p. 148-207, 2007.

CONFÉRENCE MONDIALE SUR L'ENSEIGNEMENT SUPÉRIEUR. Debats thématiques: Du traditionnel au virtuel - les nouvelles technologies de l'information. In: **L'enseignement supérieur au XXIe siècle – vision e actions**. v. 4. Paris, França: UNESCO, 1998.

COURTRIGHT, Christina. The challenge of context in information behavior research. **Annual Review of Information Science and Technology (ARIST)**, v. 41, 2006.

DERVIN, Brenda; NILAN, Michael. Information needs and uses. **Annual Review of Information Science and Technology (ARIST)**, v. 21, p. 3-33, 1986.

FAYYAD, Usama M. **Information visualization in data mining and knowledge discovery**. 1996. 346 p. Disponível em:

<<http://books.google.com.br/books?id=rYFvnyPRwkgC&printsec=frontcover&dq=fayyad&ei=joDsSZrRMY-gkQTI4Y2aAQ&client=firefox-a#PPA3,M1>>. Acesso em 14/04/2010.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. **Data mining: um guia prático**. Rio de Janeiro: Elsevier, 2005.

LE COADIC, Y. F. **A ciência da informação**. 2. ed. Brasília: Briquet de Lemos, 2004.

LITTO, F. M.; FORMIGA, M. M. M. (orgs.). **Educação a distância: o estado da arte**. São Paulo: Pearson do Brasil, 2009.

MOORE, M; KEARSLEY, G. **Educação a distância: uma visão integrada**. São Paulo: Thomson, 2007. p. 77-106.

SANCHEZ, F. (coord.). **Anuário brasileiro estatístico de educação aberta e a distância**. 4. ed. São Paulo: Instituto Monitor, 2008.

SARACEVIC, T. Interdisciplinary nature of information science. **Ciência da informação**, v. 24, n. 1, 1995.

STAIR, Ralph M. **Princípios de sistemas de informação**. 6. ed. São Paulo: Thomson, 2005.

VERGARA, Sylvia Constant. **Projetos e relatórios de pesquisa em administração**. 11. ed. São Paulo: Atlas, 2009.