

**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE - UFCG**  
**CENTRO DE CIÊNCIAS E TECNOLOGIA - CCT**  
**COORDENAÇÃO DE PÓS-GRADUAÇÃO EM INFORMÁTICA - COPIN**

**DISSERTAÇÃO DE MESTRADO**

**UM SISTEMA DE APOIO À TOMADA DE DECISÃO NO  
ACOMPANHAMENTO DO APRENDIZADO EM EDUCAÇÃO A  
DISTÂNCIA**

**Claudivan Cruz Lopes**  
**(Mestrando)**

**Ulrich Schiel**  
**(Orientador)**

**CAMPINA GRANDE, JULHO DE 2003**

**Claudian Cruz Lopes**

**Um Sistema de Apoio à Tomada de Decisão no  
Acompanhamento do Aprendizado em Educação a  
Distância**

Dissertação submetida ao Curso de Pós-Graduação em Informática do Centro de Ciências e Tecnologia da Universidade Federal de Campina Grande, como requisito parcial para a obtenção de grau de Mestre em Informática.

Área de concentração: **Ciência da Computação**

Linha: **Sistemas de Informação e Banco de Dados**

**Ulrich Schiel**

Orientador

Campina Grande, Julho de 2003

LOPES, Claudivan Cruz

L864S

Um Sistema de Apoio à Tomada de Decisão no Acompanhamento do Aprendizado em Educação a Distância.

Dissertação de Mestrado, Universidade Federal de Campina Grande, Centro de Ciência e Tecnologia, Coordenação de Pós-Graduação em Informática, Campina Grande, Paraíba, Julho de 2003.

123p. II.

Orientador: Ulrich Schiel

1. Banco de Dados
2. Mineração de Dados
3. Informática Aplicada à Educação
4. Acompanhamento do Aprendizado

CDU – 681.3.07B

**“UM SISTEMA DE APOIO À TOMADA DE DECISÃO NO  
ACOMPANHAMENTO DO APRENDIZADO EM EDUCAÇÃO A  
DISTÂNCIA”**

**CLAUDIVAN CRUZ LOPES**

**DISSERTAÇÃO APROVADA EM 29.07.2003**



**PROF. ULRICH SCHIEL, Dr.**  
**Orientador**



**MARCUS COSTA SAMPAIO, Dr.**  
**Examinador**



**PROF. ARTURO HERNANDEZ DOMÍNGUEZ, Dr.**  
**Examinador**

**CAMPINA GRANDE – PB**

*“A chave para se ter sucesso nos negócios é ter  
informações que ninguém mais tem!”*

*Aristóteles Onassis*

## AGRADECIMENTOS

---

Primeiramente a **Deus**, fonte maior de amor, misericórdia e fidelidade.

À minha filha **Yalli** e minha esposa **Priscilla**, fontes de carinho e aconchego.

Aos meus pais, **Ivan** (*In Memoriam*) e **Maria do Carmo**, e meu irmão **Carlivan**, por todo o apoio em todos os momentos de minha vida e pelo amor e dedicação que nunca poderei recompensar.

A todos os meus **familiares**, por me incentivar a ingressar no mestrado e por dar apoio em todos os momentos difíceis.

Ao meu orientador **Ulrich**, pela amizade, compreensão, tranquilidade e conhecimentos transmitidos.

Aos meus amigos, **Philip, Alex, Petrônio, Éder, Plácido, Ricardo, Wesley e Guga**, pelo apoio e incentivo, pelos ótimos momentos juntos no DSC e pelas amizades cultivadas.

Aos **professores** da COPIN, pela formação proporcionada.

Ao professor **Cláudio Baptista**, pela amizade, motivação e excelentes bate-papos.

Às funcionárias da COPIN, **Aninha** e **Vera**, que sempre resolvem nossos problemas burocráticos no mestrado.

A **CAPES**, pelo subsídio financeiro.

Na Educação a Distância - EAD, o acompanhamento do aprendizado é dificultado pela falta de contato corpo a corpo entre professores e alunos e pela falta de uma específica prática pedagógica de acompanhamento para tal modalidade de ensino. Assim, em EAD, é necessário que novos modelos ou estratégias que representem o status de aprendizagem do estudante remoto sejam elaborados. Este trabalho propõe uma estratégia de acompanhamento do aprendizado voltado para EAD baseada nas teorias de acompanhamento do ensino presencial acrescido da tática de análise de dados, onde os fatores de acompanhamento podem ser relacionados para se verificar a aprendizagem de forma mais elaborada; e apresenta o Midas-Poeta, um sistema de apoio à decisão, baseado em Mineração de Dados, para o ambiente Portfolio-Tutor. O objetivo de tal sistema é prover meios para que se possa obter padrões que caracterizem o comportamento dos alunos.

## ABSTRACT

---

In Distance Education - DE, the learning accompaniment is aggravated by the lack of body-to-body contact between teachers and students and by the lack of a specific pedagogical practice of accompaniment for such a teaching modality. This way, in DE, it is necessary to elaborate new models and strategies which represent the learning status of the remote student. This document proposes a learning accompaniment strategy suitable for DE based on accompaniment theories of presencial teaching added with data analysis tactics, where the accompaniment factors can be related together in order to verify the learning process in a more elaborated form. It presents the Midas-Poeta decision support system, using Data Mining algorithms, for the Portfolio-Tutor environment. The goal of such a system is to provide standards which characterize student's behavior.



<b>Capítulo 1. INTRODUÇÃO .....</b>	<b>1</b>
1.1. Considerações Iniciais .....	1
1.2. Motivação .....	3
1.3. Objetivos .....	4
1.4. Estrutura da Dissertação.....	4
<b>Capítulo 2. DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS E MINERAÇÃO DE DADOS .....</b>	<b>6</b>
2.1. Motivação .....	6
2.2. Descoberta de Conhecimento em Banco de Dados.....	7
2.3. Mineração de Dados .....	11
2.3.1. Tarefas de Mineração de Dados .....	11
2.3.1.1. Caracterização .....	11
2.3.1.2. Classificação.....	12
2.3.1.3. Associação.....	14
2.3.1.4. Agrupamento .....	16
2.3.2. Classificação de Sistemas de Mineração de Dados .....	17
2.3.3. Primitivas de Mineração de Dados .....	18
2.3.4. Arquiteturas de Sistemas de Mineração de Dados.....	19
2.4. Considerações Finais .....	20
<b>Capítulo 3. ACOMPANHAMENTO DO APRENDIZADO .....</b>	<b>21</b>
3.1. Avaliação da Aprendizagem.....	21
3.2. Acompanhamento do Aprendizado.....	24
3.2.1. Análise de Dados.....	28
3.3. Acompanhamento do Aprendizado na Educação a Distância - EAD.....	31
3.3.1. Educação a Distância.....	32
3.3.2. Desafios da EAD.....	34
3.3.3. Acompanhamento do Aprendizado em EAD .....	35
3.3.4. Ambientes Existentes .....	39
3.3.5. Análise dos Resultados.....	42
3.4. Considerações Finais .....	44
<b>Capítulo 4. UMA ESTRATÉGIA PARA APLICAR MINERAÇÃO DE DADOS NO ACOMPANHAMENTO DO APRENDIZADO EM EAD.....</b>	<b>45</b>
4.1. Considerações Iniciais .....	45
4.2. Requisitos da Estratégia.....	48
4.3. Descrição da Estratégia .....	50
4.3.1. Fase de Planejamento .....	50
4.3.2. Fase de Análise de Dados .....	57
4.4. Considerações Finais .....	62
<b>Capítulo 5. MIDAS-POETA: UM SISTEMA DE APOIO À TOMADA DE DECISÃO PARA O AMBIENTE PORTFOLIO-TUTOR .....</b>	<b>63</b>
5.1. O Sistema Portfolio-Tutor .....	63
5.2. O Sistema Midas-Poeta.....	70
5.2.1. Análise do Midas-Poeta .....	71
5.2.1.1. Requisitos Funcionais .....	71
5.2.1.2. Requisitos Não Funcionais .....	72
5.2.1.3. Identificação dos Atores do Sistema .....	72
5.2.1.4. Diagrama de Use-case.....	73
5.2.2. Projeto do Sistema Midas-Poeta.....	75
5.2.2.1. Fase de Planejamento do Midas-Poeta .....	76
5.2.2.2. Arquitetura do Sistema Midas-Poeta.....	80

5.2.3. Implementação do Midas-Poeta .....	84
5.2.3.1. Tecnologias Empregadas.....	84
5.2.3.2. Implementação da Fase de Análise de Dados.....	86
5.2.3.3. Disposição Física dos Módulos do Sistema.....	86
5.2.3.4. As Interfaces do Midas-Poeta.....	87
5.3. Considerações Finais .....	90
<b>Capítulo 6. UM CENÁRIO DE EXECUÇÃO DO MIDAS-POETA .....</b>	<b>92</b>
6.1. Considerações Iniciais .....	92
6.2. Interagindo com o Midas-Poeta.....	93
6.2.1. A Interação de um Professor com o Midas-Poeta.....	94
6.2.2. A Interação de um Diretor com o Midas-Poeta.....	97
6.3. Considerações Finais .....	99
<b>Capítulo 7. CONCLUSÃO .....</b>	<b>101</b>
7.1. Contribuições .....	103
7.2. Trabalhos Futuros .....	103
<b>Referências Bibliográficas .....</b>	<b>105</b>
<b>Anexo 1. TABELA DE DISCRETIZAÇÃO.....</b>	<b>111</b>

## LISTA DE FIGURAS

---

Figura 1.1. Coleção de Dados do processo de D3M.....	3
Figura 2.1. Etapas do Processo de KDD (Cabena, et al., 1998) .....	8
Figura 2.2. Etapa de Mineração de Dados ou Garimpagem.....	10
Figura 2.3. Classificador Gerado a Partir dos Dados da Instituição Financeira.....	13
Figura 2.4. Árvore de Decisão Gerada a Partir dos Dados da Instituição Financeira .....	13
Figura 2.5. Regra de Associação Gerada a Partir dos Dados da Empresa Fictícia .....	15
Figura 2.6. Definição Formal de Suporte e Confiança.....	15
Figura 2.7. Agrupamentos por Idade e Aproveitamento (Silva, 2002) .....	16
Figura 2.8. Representação Textual dos Agrupamentos.....	17
Figura 2.9. Interpretação dos Agrupamentos .....	17
Figura 4.1. Diagrama de Atividades da Fase de Planejamento.....	57
Figura 4.2. Exemplo de uma Análise de Dados (Silva, 2002).....	61
Figura 4.3. Diagrama de Atividades da Fase de Análise de Dados.....	62
Figura 5.1. Distribuição dos Alunos de uma Aula Virtual Adaptativa (Nascimento, 2002).....	66
Figura 5.2. Banco de Dados do Portfolio-Tutor (Parte Usuário) .....	66
Figura 5.3. Banco de Dados do Portfolio-Tutor (Parte Estrutura) .....	67
Figura 5.4. Banco de Dados do Portfolio-Tutor (Parte Portfolio).....	68
Figura 5.5. Banco de Dados do Portfolio-Tutor (Tutor) .....	68
Figura 5.6. Arquitetura de Alto Nível do Midas-Poeta .....	70
Figura 5.7. Diagrama de Use-Case do Sistema Midas-Poeta .....	73
Figura 5.8. Alteração do Esquema Lógico do Banco de Dados do Portfolio-Tutor.....	78
Figura 5.9. Modelo Analítico de Dados do Midas-Poeta .....	79
Figura 5.10. Conjunto de Visões do Modelo Analítico de Dados .....	79
Figura 5.11. Arquitetura do Midas-Poeta.....	81
Figura 5.12. Diagrama de Classes da Arquitetura do Midas-Poeta.....	82
Figura 5.13. Disposição Física em Três Camadas .....	86
Figura 5.14. Interface de Logon .....	87
Figura 5.15. Interfaces dos Menus de Consulta Disponível para o Professor e para o Diretor.....	87
Figura 5.16. Tela do Padrão de Consulta “Por Disciplina” .....	88
Figura 5.17. Interface do Padrão de Consulta “Por Curso” .....	88
Figura 5.18. Interface de Coleta de Dados .....	89
Figura 5.19. Interface da Etapa de Mineração.....	90
Figura 5.20. Interface da Etapa de Avaliação e Visualização dos Resultados.....	90
Figura 6.1. Etapa de Coleta de Dados Realizada pelo Professor .....	95
Figura 6.2. Resultado Mineração por Associação feita pelo Professor .....	96
Figura 6.3. Etapa de Coleta de Dados Realizada pelo Diretor.....	97
Figura 6.4. Resultado Mineração por Classificação feita pelo Diretor .....	99
Figura 7.1. Integração do sistema Midas-Poeta com um sistema Tutor.....	103

## LISTA DE TABELAS

---

Tabela 2.1. Subconjunto do Histórico de Empréstimos de Clientes (Carvalho, 2000) .....	12
Tabela 2.2. Subconjunto do Histórico de Vendas de uma Empresa Fictícia .....	14
Tabela 2.3. Dados da Turma do 3º Ano de uma Escola do Ensino Fundamental (Silva, 2002) .....	16
Tabela 3.1. Análise Comparativa dos Resultados .....	43
Tabela 4.1. Atividades Propostas e suas Validações .....	55
Tabela 4.2. Informações Disponíveis no MAD-Exemplo .....	59
Tabela 5.1. Exemplo de um Resultado Parcial do SQL Dinâmico .....	83
Tabela 5.2. Tabela de Mapeamento para Discretização .....	84
Tabela 6.1. Resultado da Coleta de Dados feita pelo Professor .....	96
Tabela 6.2. Resultado da Coleta de Dados feita pelo Diretor .....	98

## INTRODUÇÃO

Este capítulo apresenta as motivações básicas que direcionaram o desenvolvimento deste trabalho e os seus objetivos. Inicialmente faremos algumas considerações básicas a cerca do tema Avaliação e Acompanhamento do Aprendizado. Em seguida, levantaremos os problemas do acompanhamento do aprendizado em Educação a Distância e proporemos uma solução de análise de dados para favorecer este acompanhamento. Por fim, uma descrição da estrutura da dissertação será apresentada.

### 1.1. Considerações Iniciais

Ao longo da história da educação, inúmeras teorias de avaliação foram criadas com o intuito de não somente medir o desempenho do aluno, mas também como forma de caracterizá-lo e acompanhá-lo segundo o desenvolvimento de suas habilidades e conhecimentos adquiridos. Bloom *et al.* (Bloom, *et al.*, 1971) definem uma clássica taxonomia para as avaliações que surgiram até então: *avaliação formativa*, *avaliação diagnóstica* e *avaliação somativa*. É formativa toda a avaliação que ajuda o aluno a aprender e a se desenvolver, que participa da regulação das aprendizagens e do desenvolvimento no sentido de um projeto educativo. A avaliação diagnóstica tem como base o desenvolvimento de habilidades de um aluno durante um curso, onde ao seu final, é feita uma análise de seu rendimento e grau de aprendizagem. Por sua vez, a avaliação somativa é aquela baseada em testes, pontuando os alunos e servindo de julgamento para classificação dos mesmos.

No início da década de 90, Archbald e Newman (Archbald & Newman, 1989) e Wiggins (Wiggins, 1990) propuseram uma nova teoria de avaliação. Eles a chamaram de *avaliação autêntica*, na qual as melhores habilidades e capacidades dos alunos são colocadas numa ferramenta pedagógica chamada de *portfolio*, onde constam os produtos gerados ao longo do tempo pelo aluno, refletindo as suas melhores

características. Dessa forma, a avaliação autêntica proporciona além de uma auto-reflexão por parte do aluno, a verificação da aquisição de conhecimento e habilidades que o mesmo aprendeu a construir.

Independente do(s) método(s) que se utiliza(m), a avaliação tem duas atribuições básicas: a função de acompanhar o processo de aprendizagem, através de instrumentos de avaliação como provas, exercícios e testes; e a função de selecionar e classificar os estudantes, cuja finalidade é a de verificar se os alunos manifestam os comportamentos considerados pré-requisito para a aprendizagem mais avançada.

Quando planeja o ensino, o professor fixa os objetivos que pretende alcançar, que são os conhecimentos que os alunos devem adquirir ou formar durante o processo da aprendizagem. Sendo assim, é de responsabilidade do professor utilizar os meios adequados de avaliação para garantir o acompanhamento do aprendizado. Somente através do acompanhamento é que o mesmo tem a possibilidade de diagnosticar problemas de ensino/aprendizagem e executar ações que favoreçam a construção do conhecimento por parte dos alunos.

No ensino presencial, o professor pode utilizar-se de uma série de instrumentos de acompanhamento que garantirão a verificação do aproveitamento, aptidões e ajustamento pessoal-social de seus alunos. Dentre eles podemos citar os testes padronizados pela escola, cujo objetivo é avaliar o aproveitamento geral em uma ou mais disciplinas; os testes preparados pelo professor, com a função de medir a consecução de objetivos específicos de instrução; os testes de desempenho, para medir a consecução de certos objetivos que não se podem medir através de testes elaborados pelo professor nem através de teste padronizados; e as técnicas de observação de sala de aula, como uma apreciação dos aspectos do desenvolvimento pessoal que não podem ser medidas através de outros instrumentos avaliativos.

Aliada a esta base de informações, uma tendência atual no acompanhamento de aprendizagem é o processo de análise de dados, ou do inglês, *Data-driven Decision Making – D3M*, onde diversos fatores são correlacionados para se verificar a aprendizagem de maneira mais elaborada. O processo de D3M admite o uso de um banco de dados baseado no aluno, agregando informações sobre a vida escolar do estudante. Este banco de dados identificaria quem cada estudante é (dados demográficos e comportamentais), o que eles estão percebendo sobre o ambiente de estudo (dados

perceptivos) e o que eles sabem (dados sobre avaliações) (Figura 1.1). Assim, o professor pode verificar informações individuais, sumarizar os resultados de todos os estudantes (agregação) e reorganizar as informações para entender resultados sob a óptica de diferentes grupos de estudantes (desagregação e análise das informações, fazendo cruzamentos de dados).

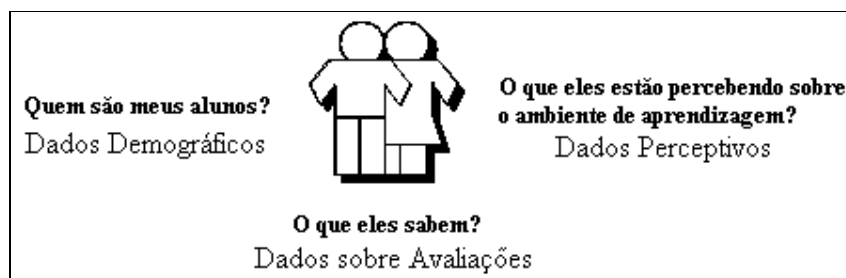


Figura 1.1. Coleção de Dados do processo de D3M

Resumindo, professores, pedagogos e administradores escolares têm uma grande base de informações que poderiam ser úteis para favorecer o acompanhamento do aprendizado, detectar problemas de ensino e tomar decisões.

## 1.2. Motivação

Na Educação a Distância – EAD, a avaliação e o acompanhamento do aprendizado são os principais problemas de sua implementação. As principais causas disso são a falta de contato corpo-a-corpo entre professores e alunos e a falta de clareza de uma prática pedagógica de acompanhamento voltada para EAD (Silva, 2002), (Menezes, *et al.*, 1999), (Silva & Fernandes, 2000), (Souto, *et al.*, 2001), (Vieira, 2002), (Nurmela, *et al.*, 1999) e outros.

Alguns trabalhos voltados para a implementação da Educação a Distância têm investigado como se pode realizar o acompanhamento de alunos em cursos à distância. Alguns deles se destinam ao acompanhamento informal, onde as interações do aluno com o ambiente de EAD são analisados de forma a acompanhar o grau de motivação e participação como na observação de sala de aula no ensino presencial. Outros estudos se destinam ao desenvolvimento de um acompanhamento alternativo aplicando avaliação autêntica e formativa em educação à distância. Porém, nenhum trabalho se destina a

aplicação de fatores informais e alternativos simultaneamente, nem aplicam o processo de D3M, proporcionando um acompanhamento mais elaborado, podendo auxiliar o professor em suas decisões pedagógicas mais efetivamente.

Assim, em EAD, é necessário que as estruturas de acompanhamento sejam aperfeiçoadas e adaptadas, e que novos modelos ou estratégias que representem o *status* de aprendizagem do estudante remoto sejam elaborados.

### 1.3. Objetivos

Os objetivos do trabalho aqui proposto são:

- Desenvolver uma estratégia de acompanhamento do aprendizado em EAD, englobando os aspectos passíveis de análise: acompanhamento informal e alternativo;
- Incrementar a estratégia de acompanhamento com a idéia de D3M, fazendo uso de dados demográficos, dados sobre avaliação, dados comportamentais, além de dados do acompanhamento informal e alternativo, onde a partir da análise destes dados, o professor poderá tomar decisões pedagógicas;
- Estruturar a estratégia com técnicas de *Data Mining* para extrair padrões de comportamentos dos alunos e favorecer o acompanhamento do aprendizado de maneira mais refinada do que as análises atualmente realizadas nos diversos ambientes de EAD. Para tal, o usuário poderá usar os dados acima descritos, armazenados ao longo do tempo e prontos para análises prospectivas, correlacionando-os sem nenhuma restrição – ou seja, ele decidirá quais dados deseja utilizar para uma mineração em questão;
- Desenvolver um sistema de apoio à decisão para o ambiente Portfolio-Tutor (Nascimento, 2002), utilizando a estratégia aqui desenvolvida.

### 1.4. Estrutura da Dissertação

Esta dissertação é constituída de sete capítulos, incluindo esta introdução. Os capítulos restantes são organizados da seguinte forma:



- O capítulo 2 apresenta uma revisão bibliográfica sobre Mineração de Dados, cobrindo os aspectos relacionados ao tema que aqui nos interessa;
- O capítulo 3 apresenta uma revisão bibliográfica sobre o acompanhamento do aprendiz, a problemática do acompanhamento do aprendiz em EAD e os trabalhos relacionados;
- No capítulo 4 detalharemos a estratégia de acompanhamento do aprendiz aqui proposta;
- O capítulo 5 mostra os aspectos de análise, projeto e implementação do Midas-Poeta, um sistema de apoio à tomada de decisão para o ambiente Portfolio-Tutor. Tal sistema será concebido segundo a estratégia de acompanhamento descrita no capítulo 4;
- No capítulo 6 apresentaremos um estudo de caso mostrando as potencialidades do sistema Midas-Poeta;
- Por fim, no capítulo 7 faremos nossas conclusões sobre o trabalho desenvolvido e apresentaremos as contribuições do mesmo.

### *DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS E MINERAÇÃO DE DADOS*

Para fundamentar os processos de Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases* – KDD) e Mineração de Dados (*Data Mining* – DM), é importante diferenciar os conceitos de dado, informação e conhecimento.

Setzer (Setzer, 2001) define dado como uma seqüência de símbolos quantificados ou quantificáveis e que não têm significado além de sua existência. Já a informação é uma abstração informal que dá significado a um conjunto de dados. Enquanto que conhecimento é um conjunto apropriado de informações úteis e valiosas. Dessa forma, existe informação que se relaciona a um conhecimento, mas pode haver informação sem essa relação.

Considerando que os bancos de dados armazenam grandes quantidades de dados, presume-se que uma grande quantidade de informações pode ser extraída. Neste capítulo, apresentaremos os processos de KDD e DM como ferramentas utilizadas para, a partir dos dados numa base de dados, extrair informações desconhecidas e potencialmente úteis, de modo que estas possam representar um conhecimento.

#### 2.1. Motivação

Ao longo da história da humanidade, o conhecimento tem sido o fator decisivo que diferencia as pessoas e os negócios. Aquele que o busca, se sobressai.

Aliada à busca pelo conhecimento, está a extraordinária quantidade de dados de que dispomos atualmente. Estes dados têm sido armazenados em grandes bancos de dados os quais vêm crescendo assustadoramente graças à evolução das tecnologias atuais. No entanto, a medida em que mais dados são colocados a nossa disposição, maior é a lacuna entre a geração destes dados e o entendimento deles. Portanto, um

problema que se posa é: como processar esta magnitude de dados transformando-os em informações úteis?

Além da forma tradicional, através do uso de linguagens de consultas como SQL, são necessárias eficientes técnicas e ferramentas de recuperação da informação capazes de apoiar as decisões de negócios tomadas por usuários mais sofisticados. É neste contexto que surge a Descoberta de Conhecimento em Bases de Dados como um processo que envolve a automação parcial da identificação e do reconhecimento de padrões em uma grande base de dados (Adriaans & Zantigue, 1996), (Fayyad, 1998).

## 2.2. Descoberta de Conhecimento em Banco de Dados

KDD é um campo multidisciplinar e integra métodos de diversas áreas incluindo estatística, banco de dados, inteligência artificial, aprendizagem automática, visualização de dados, reconhecimento de padrões, otimização e performance computacional e outros.

De acordo com Cabena *et al.* (Cabena, *et al.*, 1998), descobrir conhecimento em bancos de dados significa extrair, de grandes bases de dados, informações desconhecidas, válidas e acionáveis, que podem ser utilizadas para a tomada de decisão.

Informação desconhecida significa geração de um conhecimento totalmente novo. A busca por informações válidas está intimamente ligada aos objetivos do negócio em questão. Portanto, é necessária uma compreensão da informação gerada a fim de extrair o conhecimento realmente relevante. Por fim, a informação deve ser acionável, pois o conhecimento produzido deve ser traduzido em ações que favoreçam ao crescimento da empresa, pessoa ou negócio.

A fim de se obter informação desconhecida, útil, compreensível, válida e acionável, deve-se utilizar um processo de descoberta de conhecimento. De acordo com Cabena *et al.* (Cabena, *et al.*, 1998), KDD como um processo é descrito na Figura 2.1 e consiste de uma seqüência interativa e iterativa das seguintes etapas: **Seleção, Pré-processamento, Transformação, Garimpagem e Análise e Assimilação dos Resultados.**

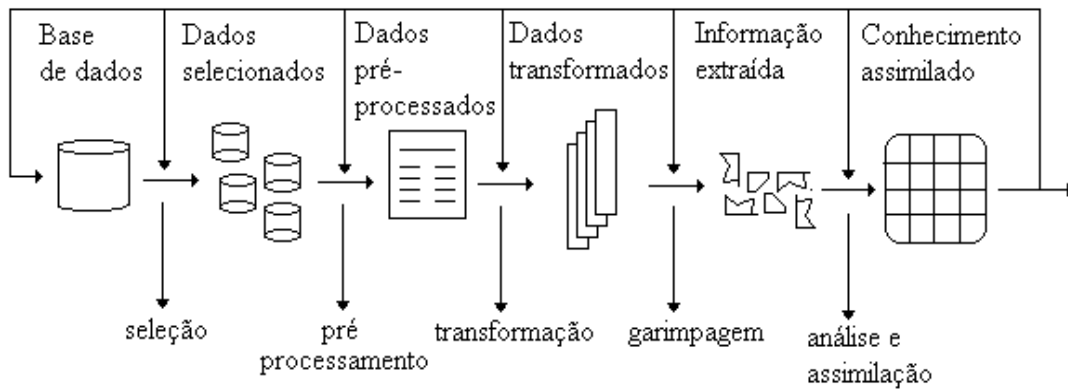


Figura 2.1. Etapas do Processo de *KDD* (Cabena, *et al.*, 1998)

Embora não ilustrado na Figura 2.1, KDD começa com o entendimento e definição do domínio da aplicação e dos objetivos finais a serem atingidos. Para isso, é necessária a presença de um analista de negócio, que possua todo o entendimento sobre o domínio em questão. Em seguida, as atividades de Seleção, Pré-processamento e Transformação compõem o que se chama de Preparação de Dados. Seguindo no processo, chegamos a fase de Garimpagem ou Mineração de Dados, cuja finalidade é aplicar algoritmos de descoberta de conhecimento aos dados preparados produzindo modelos de conhecimento. Por fim, temos a etapa de Análise e Assimilação de Conhecimento, cujo objetivo é interpretar os padrões gerados e utilizá-los para dar suporte à tomada de decisão.

Na etapa de Seleção, são identificados as bases de dados e quais atributos e exemplos de dados devem ser extraídos para a fase de Mineração de Dados.

A fase de Pré-processamento destina-se a melhorar a qualidade dos dados e conseqüentemente os resultados da mineração, pois os dados selecionados na etapa anterior podem estar inconsistentes e incompletos. Dentre as principais rotinas de pré-processamento, podemos destacar:

- **Limpeza:** são rotinas aplicadas para preencher dados faltantes, corrigir dados ‘sujos’ e inconsistências. Dados faltantes são aqueles que de fato existem, mas que por algum motivo não estão presentes no banco de dados (Toebe, 2002). Por exemplo, em um cadastro de clientes, o campo *data\_de\_nascimento* pode não ter sido preenchido para alguns clientes. Dados ‘sujos’ são informações corrompidas que causam inconsistências no banco de dados. As causas de dados

‘sujos’ são várias. Os exemplos mais comuns são oriundos da fase de entrada de dados como erros de digitação;

- **Redução:** são técnicas aplicadas ao conjunto de dados no sentido de reduzir o tamanho dos dados. A redução dar-se-á por agregação, onde operações de agregação são aplicadas aos dados; por agrupamento, onde informações similares são organizadas em grupos; e por generalização, onde atributos de baixo nível são substituídos por conceitos de alto nível. Por exemplo, num cadastro de clientes, considere uma variável chamada *cidade*. Este poderia ser substituído por um conceito de alto nível como estado ou região;
- **Discretização:** técnica utilizada para transformar um domínio de dados de valores contínuos em valores discretos. Por exemplo, considere uma variável contínua no banco de dados que armazene a idade dos clientes. Pode-se transformar essa variável contínua em variável discreta utilizando faixas de valores como jovem, adulto, maduro e idoso;
- **Amostragem:** é o processo de selecionar uma amostra representativa da população de dados. A importância desta rotina está no alto custo de processamento dos algoritmos de mineração se todos os dados disponíveis fossem processados (Pyle, 1999), (Brumem, *et al.*, 2001);
- **Fragmentação:** são técnicas utilizadas para a separação dos dados em dois conjuntos: o conjunto de treinamento e conjunto de testes. O conjunto de treinamento é utilizado como entrada para um algoritmo de mineração inferir padrões de conhecimento. O conjunto de testes é utilizado para verificar a acurácia dos padrões encontrados (Kohavi, 1995), (Witten & Frank, 2000), (Jain & Dubes, 1997).

Para finalizar a preparação de dados, o último passo é a Transformação de Dados, cuja finalidade é adequar os dados selecionados e pré-processados de acordo com a entrada específica de cada algoritmos de mineração de dados.

Prosseguindo no processo, chega-se a etapa de Mineração de Dados ou Garimpagem, que começa com a escolha dos algoritmos a serem aplicados. Essa escolha depende fundamentalmente do objetivo do processo de KDD. Entre os

principais métodos existentes pode-se citar regras de associação, classificação, agrupamento, regressão, etc.

De modo geral, na fase de Mineração de Dados, algoritmos especializados procuram padrões nos dados. Esta busca pode ser efetuada automaticamente pelo sistema ou interativamente com um analista, responsável pela geração das hipóteses. Porém, em geral, o processo de busca é interativo, de forma que os analistas possam rever os resultados e formar um novo conjunto de questões para refinar a busca em um dado aspecto das descobertas e realimentar o sistema com novos parâmetros.

A Figura 2.2 abaixo ilustra a etapa de Mineração de Dados. O algoritmo de *Data Mining* recebe como entrada um conjunto de dados preparado e gera um Modelo ou Esquema de Conhecimento como saída. Para cada tarefa de Mineração, temos um modelo distinto (regras, árvores de decisão, lista de decisão, etc).

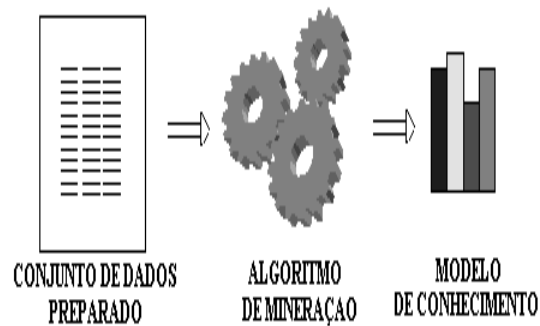


Figura 2.2. Etapa de Mineração de Dados ou Garimpagem

Ao final do processo, a Fase de Pós-Processamento (Fayyad, 1998) ou de Análise e Assimilação dos Resultados, irá gerar relatórios, modelos ou representações dos padrões descobertos, que passam então a ser interpretados e validados pelo especialista do domínio, de forma que o modelo produzido seja acionável e relevante para o negócio em questão. Caso a avaliação seja positiva, então ocorre a incorporação ou assimilação do conhecimento gerado. Do contrário, todo ou parte do processo de KDD deve ser revisado.

## 2.3. Mineração de Dados

A mineração de dados é a etapa central no processo de KDD que executa a descoberta de conhecimento propriamente dita, pois são os algoritmos de mineração de dados que de forma semi-automática produzem o conhecimento a partir de dados existentes. Nesta seção, trataremos dos principais aspectos relacionados a DM.

### 2.3.1. Tarefas de Mineração de Dados

As tarefas de DM dizem respeito à aplicação de um algoritmo de mineração e ao tipo de padrão que pode ser minerado. A seguir, apresentaremos as principais tarefas de mineração.

#### 2.3.1.1. Caracterização

Uma caracterização de dados é a identificação de características gerais entre instâncias de uma certa classe de dados. Por exemplo, um sistema de DM para o setor de vendas de uma empresa constatou que dos clientes que gastam mais que R\$10.000,00 por ano, 90% deles são entre 40 e 50 anos, tem emprego e possuem uma excelente taxa de crédito. Portanto, os clientes com este perfil são caracterizados como ótimos clientes, têm descontos especiais e tratamento diferenciado, como prazos e condições de pagamento.

Uma técnica utilizada para a tarefa de caracterização é o método de *AOI* (*Attribute-Oriented Induction*), em português, Indução Orientada aos Atributos (Cai, *et al.*, 1991), (Han, *et al.*, 1993), (Han & Fu, 1996), (Carter & Hamilton, 1998), (Han, *et al.*, 1998). Esta técnica realiza a sumarização examinando o número de valores distintos de cada atributo no conjunto de dados. Agregações são realizadas fundindo tuplas idênticas e acumulando suas respectivas contagens. A caracterização resultante pode ser mapeada em diferentes formas de representação ao usuário, tais como gráficos ou regras.

### 2.3.1.2. Classificação

Classificação de dados é o processo de prever a classe de um objeto segundo certos atributos de predição (Han & Kamber, 2001).

Para construir o modelo de classificação, também chamado de classificador, deve-se descobrir uma relação entre os atributos de predição com o atributo de classificação.

Para isso, o conjunto de dados é fragmentado em duas amostras chamadas de conjunto de treinamento e conjunto de teste. O primeiro conjunto é utilizado para que o algoritmo de classificação descubra a relação descrita anteriormente. O segundo conjunto é utilizado para verificar a acurácia do classificador gerado. O modelo de classificação resultante é então usado para prever a classe de registros não classificados (Chen, *et al.*, 1996).

Para exemplificar esta tarefa, suponha que uma instituição financeira tem interesse em descobrir classes de clientes confiáveis (clientes cujos empréstimos apresentam um baixo risco de prejuízo) e clientes não confiáveis (cujos empréstimos seriam altamente arriscados) para classificar novos pedidos de empréstimos. Na montagem destas duas classes é necessário que exista no conjunto de treinamento um histórico sobre os empréstimos dos clientes, bem como a situação dos mesmos: pago ou não pagos. Para gerar as classes, os atributos de predição escolhidos pelo especialista foram a renda anual (discretizada) e o grau de escolaridade dos clientes. Um subconjunto do histórico de empréstimos dos clientes é ilustrado na Tabela 2.1 abaixo:

ESCOLARIDADE	RENDA ANUAL	PAGOU EMPRÉSTIMO?
Mestrado	Alta	Sim
Mestrado	Média	Não
.....	.....	.....
2º Grau	Alta	Sim
Mestrado	Baixa	Não

**Tabela 2.1.** Subconjunto do Histórico de Empréstimos de Clientes (Carvalho, 2000)

Como resultado, o algoritmo de classificação gerou o classificador apresentado na Figura 2.3 abaixo.



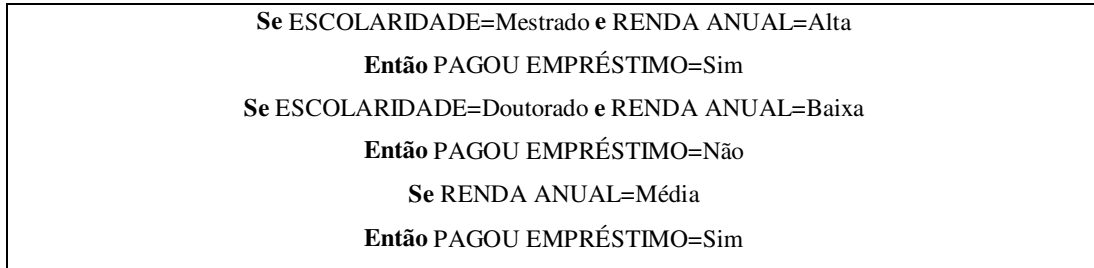


Figura 2.3. Classificador Gerado a Partir dos Dados da Instituição Financeira

Sua interpretação é a seguinte: uma pessoa que tem o nível de escolaridade igual ao mestrado e possui renda anual Alta é uma pessoa que paga seus empréstimos. Já uma pessoa que tem o nível de escolaridade igual ao doutorado e renda anual baixa não cumpre com suas obrigações. E por fim, uma pessoa que possui renda anual média sempre paga os seus empréstimos. Tal conhecimento extraído permite ao gerente tomar a decisão de fazer novos empréstimos com maior segurança.

De maneira análoga, as regras produzidas poderiam ser expressas sob a forma de uma árvore de decisão.

Árvores de decisão são meios de representar os resultados de uma classificação na forma de um gráfico organizacional. Por exemplo, o classificador da Figura 2.3 poderia ser representado através de uma árvore de decisão como segue.

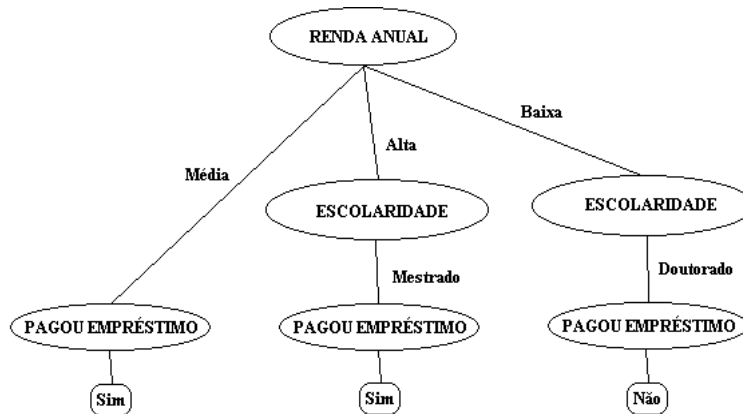


Figura 2.4. Árvore de Decisão Gerada a Partir dos Dados da Instituição Financeira

Um algoritmo clássico que implementa árvores de decisão é o *ID3* (Quinlan, 1986) e suas variantes, *C4.5* e *C5* (Quinlan, 1993), (See5, 1999). Este algoritmo varre todos os dados de treinamento e começa a construir a árvore de forma *top-down*. Inicialmente, é escolhido um atributo “a” com a melhor função de avaliação (função que

deve gerar um subconjunto de dados com a maioria dos exemplos pertencentes à mesma classe) para particionar estes dados. Para cada valor “i” do atributo “a”, um ramo “r” é criado junto com o correspondente subconjunto de dados que possuem o valor de “a” igual a “i”. Assim, para cada ramo “r”, é criado um nó “n” na árvore que poderá ser uma classe (caso todos os exemplos analisados tenham “a” igual a “i” e representem a mesma classe) ou um outro atributo “a” (se algum exemplo de “a” igual a “i” representar classes distintas) para se fazer a nova partição. O algoritmo é executado recursivamente em todos aqueles nós que não representarem uma classe.

Podemos ainda citar outros algoritmos de classificação como o *PRISM* (Cendrowska & Stutz, 1987), *OneR* (Holte, 1993) e Naive Bayes (John & Langley, 1995).

### 2.3.1.3. Associação

Uma regra de associação é definida como: *se X então Y* ou  $X \Rightarrow Y$ , onde X e Y são conjuntos de itens no formato *atributo-valor* e  $X \cap Y = \emptyset$ . Diz-se que X é o antecedente da regra e Y é o conseqüente da regra (Carvalho, 2000). A regra de associação  $X \Rightarrow Y$  é interpretada como “as tuplas da base de dados que satisfazem a condição X satisfaz também a condição Y”.

Por exemplo, considere que a Tabela 2.2 apresenta as vendas de uma empresa fictícia que comercializa produtos eletrônicos. Os atributos selecionados pelo especialista foram a idade do cliente, discretizada em faixa etária (jovem, adulto, maduro, idoso), a renda mensal e o produto comprado.

FAIXA ETÁRIA	RENDA MENSAL	COMPRA
JOVEM	Alta	CD PLAYER
ADULTO	Média	TV
.....	.....	.....
JOVEM	Alta	CD PLAYER
MADURO	Baixa	BATEDEIRA
JOVEM	Alta	DVD PLAYER

**Tabela 2.2.** Subconjunto do Histórico de Vendas de uma Empresa Fictícia

Uma tarefa de associação poderia gerar uma regra conforme a Figura 2.5, onde X é uma variável que representa um cliente.

$$\text{FAIXA ETÁRIA (X, "JOVEM")} \wedge \text{RENDA MENSAL (X, "ALTA")} \Rightarrow \text{COMPRA (X, "CD PLAYER")}$$

(Suporte=40%, Confiança=60%)

Figura 2.5. Regra de Associação Gerada a Partir dos Dados da Empresa Fictícia

A regra indica que dos clientes da empresa fictícia, 40% estão na faixa etária JOVEM com renda mensal ALTA e compram CD PLAYER. Este percentual é chamado de **Suporte** e indica a porcentagem de registros no banco de dados que satisfazem a regra de associação dada.

Continuando a interpretação da regra, existe uma probabilidade de 60% em que clientes nesta faixa etária e com esta renda comprem CD PLAYER. Este percentual é chamado de **Confiança** e indica o grau de acerto da regra. Formalmente, o suporte e a confiança são definidos como:

$$\text{Suporte } (X \Rightarrow Y) = (\text{n}^\circ \text{de tuplas contendo X e Y}) / (\text{n}^\circ \text{total de tuplas})$$
$$\text{Confiança } (X \Rightarrow Y) = (\text{n}^\circ \text{de tuplas contendo X e Y}) / (\text{n}^\circ \text{de tuplas contendo X})$$

Figura 2.6. Definição Formal de Suporte e Confiança

No tocante a implementação de tarefas de associação, muitos algoritmos foram desenvolvidos, dentre os quais podemos destacar: *LargeKItemSets* (Agrawal, *et al.*, 1993), *Apriori* (Agrawal & Srikant, 1994), *AprioriTid* (Agrawal & Srikant, 1994), *Partition* (Savasere, 1995) e *Multiple Level* (Srikant & Agrawal, 1996). Destes, o mais utilizado é o *Apriori*, sendo os demais, ou extensão dele, ou o utilizam.

O algoritmo *Apriori* trabalha basicamente fazendo uma varredura sobre os dados a fim de gerar todos os conjuntos de combinações de valores de colunas que aparecem nos dados. Posteriormente, são considerados apenas aqueles conjuntos com uma frequência não menor que o suporte mínimo pré-fixado, gerando o denominado **conjunto G de grandes conjuntos**. Após a geração do conjunto G de grandes conjuntos, é necessário extrair as regras de associação sobre cada elemento de G que satisfaça o fator de confiança mínima.

#### 2.3.1.4. Agrupamento

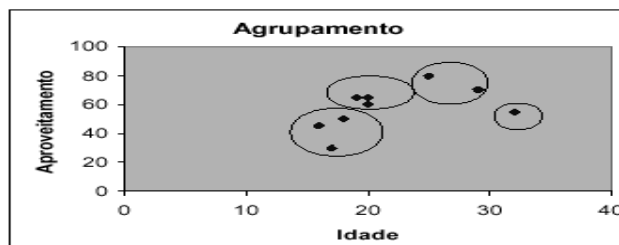
Basicamente, a tarefa de agrupamento ou *clustering* traduz-se em agrupar um conjunto de dados utilizando algum critério de similaridade e diferentemente da tarefa de classificação, estes dados não devem ter sido previamente classificados. Para diferenciar os grupos há um princípio conceitual de “máxima similaridade intraclasse e mínima similaridade interclasse” (Han & Kamber, 2001). Ou seja, grupos de objetos são formados tal que objetos dentro de um grupo têm alta similaridade, porém são muito diferentes dos objetos presentes em outro grupo.

Todo grupo que é formado pode ser visualizado como uma classe de objetos na qual regras podem ser derivadas. Por exemplo, na Tabela 2.3 são mostrados os dados de uma turma do 3º ano de uma escola do ensino fundamental

APROVEITAMENTO (%)	IDADE
45	16
65	20
65	19
55	32
70	29
80	25
30	17
60	20
50	18

**Tabela 2.3.** Dados da Turma do 3º Ano de uma Escola do Ensino Fundamental (Silva, 2002)

Neste caso, um resultado possível é o que agrupa os alunos em 4 categorias apresentados na Figura 2.7.



**Figura 2.7.** Agrupamentos por Idade e Aproveitamento (Silva, 2002)

Textualmente, o mesmo resultado poderia ser escrito da seguinte forma:

Alunos com idade entre 16 e 18 têm aproveitamento inferior a 60%
Alunos com idade entre 19 e 20 têm aproveitamento entre 60% e 65%
Alunos com idade entre 25 e 29 têm aproveitamento entre 70% e 80%
Alunos com idade maior que 29 têm aproveitamento inferior a 60%

Figura 2.8. **Representação Textual dos Agrupamentos**

A partir destas categorias, a seguinte interpretação poderia ser derivada:

<b>Se</b> $16 \leq \text{IDADE} \leq 18$ <b>ou</b> $\text{IDADE} > 29$ <b>então</b> $\text{APROVEITAMENTO} < 60\%$
<b>Se</b> $19 \leq \text{IDADE} \leq 20$ <b>então</b> $60\% \leq \text{APROVEITAMENTO} \leq 65\%$
<b>Se</b> $25 \leq \text{IDADE} \leq 29$ <b>então</b> $70\% \leq \text{APROVEITAMENTO} \leq 80\%$

Figura 2.9. **Interpretação dos Agrupamentos**

No que diz respeito à implementação da tarefa de agrupamento, diversos algoritmos foram propostos. Entre eles, podemos destacar *k-Means* (MacQueen, 1967), *k-Medoids* (Kaufman & Rousseeuw, 1990), *k-Modes* e *k-Prototypes* (Huang, 1998), *Cobweb* (Fisher, 1987) e o *Expectation Maximization - EM* (Lauritzen, 1995).

O algoritmo *k-Means* trabalha da seguinte forma: dado um banco de dados de “n” objetos e “k” grupos a formar, ele randomicamente seleciona os objetos em “k” partições, cada um representando um grupo. Para cada um dos objetos restantes, um objeto é associado ao grupo no qual ele é mais similar baseada na função de similaridade *Erro Quadrático* que dá a distância do objeto ao centro do grupo. O algoritmo então computa um novo centro para cada grupo. Este processo é iterativamente repetido até a função de similaridade convergir. Já o algoritmo *EM* estende o *k-Means* da seguinte forma: ao invés de associar cada objeto a algum grupo de acordo com a distância dele ao centro do grupo, o algoritmo associa cada objeto ao grupo de acordo com um peso representando a probabilidade de similaridade. Portanto, novos grupos são computados baseados no peso medido.

### 2.3.2. Classificação de Sistemas de Mineração de Dados

Segundo Han e Kamber (Han & Kamber, 2001), um sistema de DM pode ser categorizado segundo vários critérios como segue:

- **Tipo de Banco de Dados Minerado:** sistemas de DM podem ser classificados de acordo com o modelo de dados (relacional, orientado a objetos, objeto-relacional ou multidimensional) ou de acordo com os tipos de dados a serem minerados (dados espaciais, multimídia, textuais, etc)
- **Tipo de Conhecimento Minerado:** sistemas de DM podem ser classificados de acordo com o tipo de conhecimento minerado (regras de associação, regras de classificação, caracterização, agrupamento, etc);
- **Tipo de Técnica Utilizada:** a técnica utilizada descreve o grau de interação do usuário envolvido (minerador direcionado por consultas, minerador interativo e exploratório);
- **Tipo de Aplicação Adotada:** sistemas de DM podem ser classificados segundo a aplicação adotada (finanças, educação, fraudes, vendas, etc).

### 2.3.3. Primitivas de Mineração de Dados

Uma concepção popular errônea sobre mineração de dados é que um sistema de DM pode autonomamente descobrir conhecimento sem a intervenção humana (Han & Kamber, 2001). Na prática, tal sistema poderia gerar um padrão incompreensível ou um enorme conjunto de regras, o que seria impraticável.

Um cenário mais realista é supor que os usuários possam interagir com o sistema utilizando um conjunto de primitivas de DM desenvolvidas para facilitar a descoberta de conhecimento e guiar o processo de DM, tornando-o um processo de DM exploratório. Tais primitivas incluem:

- **Dados relevantes à tarefa de mineração:** tipicamente os usuários estão interessados num subconjunto da base de dados. Por exemplo, suponha que o gerente de uma loja multinacional de produtos eletrônicos é responsável pelas regiões dos EUA e Canadá. Porém, ele deseja estudar as compras dos clientes apenas da região da Califórnia dos EUA. Ao invés de minerar toda a base de dados, o gerente pode especificar que apenas os dados relativos aos clientes da Califórnia necessitam ser recuperados para análises prospectivas. Este processo de coleta de dados resulta num novo conjunto de dados que pode ser sumarizado e pré-processado antes de aplicar a tarefa de DM;

- **Tipo de conhecimento a ser minerado:** o usuário especifica a tarefa de mineração a ser realizada tal como caracterização, associação, classificação, agrupamento etc. Além disso, o usuário também pode prover *templates* no qual todos os padrões descobertos devem segui-los. Exemplificando, suponha que o gerente da loja multinacional de produtos eletrônicos do exemplo anterior deseja estudar os hábitos de compras de seus clientes. Ele pode minerar regras de associação de acordo com o *template*  $P(X: \text{cliente}, W) \wedge Q(X, Y) \Rightarrow \text{Compra}(X, Y)$ , onde X é um cliente; P e Q são variáveis predicativas que podem ser instanciadas segundo os atributos relevantes especificados como parte dos dados relevantes a tarefa. Logo, a pesquisa de associação deve produzir regras que casem com o dado *template*, tais como  $\text{idade}(X, "30..39") \wedge \text{renda}(X, \text{Alta}) \Rightarrow \text{compra}(X, \text{"Computador"})$  e  $\text{ocupação}(C, \text{"Estudante"}) \wedge \text{idade}(X, "20..29") \Rightarrow \text{compra}(X, \text{"DVD Player"})$ ;
- **Medidas de interesse:** embora as primitivas “dados relevantes à tarefa de mineração” e “tipo de conhecimento a ser minerado” possam guiar a mineração a resultados satisfatórios, ainda assim, uma grande quantidade de padrões pode ser gerada. Tipicamente, apenas uma parte destes padrões é útil ao minerador. Logo, são necessárias medidas capazes de filtrar as regras produzidas ou para avaliar o conhecimento minerado. Dentre as medidas podemos citar simplicidade (a regras deve ser facilmente compreendida por humanos), certeza (a cada padrão descoberto deve ser associado uma medida que avalie a certeza ou confiabilidade, tal como a confiança) e utilidade (uma medida que avalie o potencial de utilidade da regra, tal como o suporte);
- **Apresentação e visualização dos padrões descobertos:** esta primitiva se refere à forma na qual os padrões descobertos são apresentados ao usuário. O usuário pode escolher diferentes formas de representação do conhecimento como regras, tabelas, gráficos, grafos, árvores de decisão, listas de decisão e outros.

#### 2.3.4. Arquiteturas de Sistemas de Mineração de Dados

Uma questão crítica no projeto de sistemas de DM é sua integração de com sistemas de banco de dados ou com sistemas de *data warehousing*. Para responder esta

questão, precisamos examinar as possíveis formas de acoplar ou integrar estas tecnologias. A seguir, apresentaremos os esquemas de integração utilizados na atualidade.

- **Sem acoplamento:** significa que um sistema de DM não utiliza nenhum recurso de um sistema de banco de dados ou de *data warehousing*. Neste caso, o sistema de DM carrega dos dados de uma fonte de dados tal como um *flat file*, processa os dados usando algum algoritmo de mineração e armazena os resultados em outro arquivo;
- **Fracamente acoplado:** significa que um sistema de DM usa os recursos de um sistema de banco de dados ou de *data warehousing* para carregar os dados de um repositório de dados gerenciado por estes sistemas. Porém, a mineração é feita utilizando ferramentas externas a tais sistemas;
- **Fortemente acoplado:** significa que o sistema de DM é integrado ao sistema de banco de dados ou de *data warehousing*. Portanto, o subsistema de DM é tratado como um componente funcional do sistema.

Com esta apresentação, podemos constatar que a abordagem fracamente acoplada é melhor que a abordagem sem acoplamento, já que a primeira faz uso dos recursos do sistema de banco de dados ou de *data warehousing*. Já a o esquema de integração fortemente acoplado é altamente desejável, porém sua implementação não é trivial e mais pesquisas nesta área precisam ser feitas.

## 2.4. Considerações Finais

Este capítulo abordou os aspectos relativos ao processo de Descoberta de Conhecimento em Bancos de Dados – KDD e Mineração de Dados - DM. Para KDD, foram definidos os conceitos e as etapas do processo. Para a Mineração de Dados, mostramos e exemplificamos as principais técnicas, apresentamos uma classificação de sistemas de DM, as primitivas de DM e as diferentes arquiteturas de sistemas de DM.



### *ACOMPANHAMENTO DO APRENDIZADO*

O objetivo deste capítulo é apresentar a avaliação e o acompanhamento do aprendizado. Inicialmente, explanaremos o papel pedagógico da avaliação dentro do processo educacional e falaremos sobre a análise de dados. Esta última, recentemente, tornando-se uma estratégia bastante utilizada pelos educadores para favorecer o acompanhamento do aprendizado. Por fim, a problemática do acompanhamento do aprendizado em Educação a Distância – EAD e os trabalhos relacionados serão mostrados.

#### 3.1. Avaliação da Aprendizagem

Um papel fundamental da educação e dos educadores é permitir a construção humana e moral das pessoas disponibilizando informações que viabilizem a formação do aprendizado e conseqüentemente a assimilação do conhecimento. Só que a grande quantidade de informações e a volatilidade destas informações passa a exigir uma postura diferente no processo de ensino e aprendizagem. Para o aluno, memorizar o conteúdo deixa de ser o principal objetivo. Aprender a buscar e utilizar o conhecimento que for preciso e no momento que for necessário, passa a ser o mais importante. Além disso, o mercado de trabalho também tem exigido outras habilidades do profissional moderno como criatividade, liderança, pensamento crítico, trabalho em equipe, atitude pró-ativa, dentre outras (Nascimento, 2002). Conseqüentemente, como os educadores podem acompanhar a construção do conhecimento por parte dos alunos? Quais áreas de conhecimento foram trabalhadas pelos alunos? Quais os avanços que os alunos vêm demonstrando? Como descobrir se as estratégias didáticas têm alcançado os objetivos traçados? As respostas destas perguntas podem ser obtidas através da avaliação e do acompanhamento da aprendizagem.

Para o aluno, a avaliação é importante pois permite ao mesmo conhecer os resultados de seu empenho e esforço, não só pela satisfação da aprendizagem mas, especialmente, pelo significado que tem o conhecimento de suas capacidades para futuras aprendizagens. Assim, a avaliação serve para os alunos como um instrumento de diagnóstico de sua situação, tendo em vista a definição de encaminhamentos adequados para a sua aprendizagem.

A avaliação é importante para o professor, pois as informações que ele vai obtendo no processo avaliativo auxiliam-no a modificar gradativamente seu parecer sobre o aluno, em relação a suas potencialidades e suas limitações, sem esquecer que cada estágio de desenvolvimento e de conhecimentos dos indivíduos é provisório. Com base na análise dessas informações e da conduta do aluno, é possível ao professor formular hipóteses sobre o próprio processo de ensino e aprendizagem e sobre as etapas de desenvolvimento de cada um dos alunos.

Mas, o que é avaliar? Qual é o seu papel no processo educacional?

De acordo com Dietel, *et al.* (Dietel, *et al.*, 1991), a avaliação pode ser definida como qualquer método para melhor entender o conhecimento que um estudante possui e deve ser instrumento auxiliar do aprendizado em uma forma construtivista de aquisição do conhecimento.

Para Hoffmann (Hoffmann, 1993), a avaliação significa ação provocativa do professor, desafiando o educando a refletir sobre as situações vividas, a formular e reformular hipóteses, encaminhando-se a um saber enriquecido.

Segundo Luckesi (Luckesi, 2000), a avaliação da aprendizagem deve ser um processo amoroso e inclusivo, acolhendo o estudante do jeito que ele é, com suas dificuldades e necessidades; e o ato de avaliar deve ser composto por duas atividades articuladas e indissociáveis: diagnosticar e decidir. O diagnóstico envolve a constatação do estado de aprendizado do aluno através da aplicação de instrumentos avaliativos como provas, seminários, exames, testes, exercícios, práticas entre outros. A decisão é a reflexão sobre qual direcionamento seguir para melhorar ou modificar as práticas didático-pedagógicas caso o aprendizado tenha ou não tenha sido satisfatório.

A avaliação deve verificar como o aluno é capaz de movimentar-se num campo de estudos e estimulá-lo, através de uma reflexão conjunta sobre o que ele realizou, a encontrar os caminhos do seu próprio desenvolvimento. Portanto, a avaliação no

processo educacional, somente tem sentido na medida em que sirva para o diagnóstico do processo de ensino-aprendizagem, em função dos resultados que estão sendo buscados na ação educativa, verificando o desenvolvimento das habilidades e acompanhando o progresso dos alunos durante todo o processo de aprendizagem (Melchior, 1999).

Em Bloom *et al.* (Bloom, *et al.*, 1971), existe uma clássica taxonomia para as formas de avaliação divididas em três tipos:

- **Avaliação Formativa:** o professor utiliza um conjunto de questões, testes e tarefas como forma de auxiliar o aluno para o seu efetivo aprendizado, fornecendo *feedback* necessário ao seu aperfeiçoamento, sem necessariamente contabilizar sua pontuação no resultado final.

Segundo Boston (Boston, 2002), ao aplicar este método de avaliação, o professor pode usar este *feedback* para fazer os ajustes instrucionais necessários tais como ensinar novamente algum assunto, tentar abordagens e metodologias instrucionais alternativas ou oferecer mais oportunidades para a prática;

- **Avaliação Somativa:** tem como função principal determinar o nível de desempenho alcançado pelos alunos durante o curso, pontuando os alunos e servindo como parâmetro de julgamento para classificação dos mesmos, normalmente privilegiando o aspecto de capacidade de memorização;
- **Avaliação Diagnóstica:** objetiva medir o grau de habilidades e de conhecimentos dos alunos sobre o assunto a ser tratado durante a realização do curso, e ao final como forma de levantar as possíveis falhas de aprendizagem.

Além destas formas de avaliação, podemos considerar a Avaliação Autêntica proposta por Archbald e Newman (Archbald & Newman, 1989) e Wiggins (Wiggins, 1990). Neste tipo de avaliação, as capacidades do aluno e os produtos gerados pelo mesmo ao longo do tempo devem oferecer:

- oportunidades para auto-reflexão por parte do aluno, pois o aluno assume cada vez mais o controle e a responsabilidade sobre o seu processo de aprendizagem;
- verificação da aquisição de conhecimento, de habilidades e atitudes comprovadas através dos resultados de todos os seus trabalhos;

- transparência, por estar disponível a todos os envolvidos no processo educacional como o aluno, o professor, a instituição e a família do aluno.

Na avaliação autêntica, é utilizada uma ampla gama de instrumentos avaliativos como desenvolvimento de projetos, pesquisas, debates, seminários, experiências de laboratórios, entrevistas, etc; e uma série de itens de avaliação subjetivos é levada em consideração no processo de julgamento. São exemplos de itens de avaliação: criatividade, capacidade de síntese, capacidade crítica, liderança, etc.

Uma estratégia da aplicação da avaliação autêntica é a utilização de portfólios, que correspondem a arquivos ou pastas que armazenam um conjunto de trabalhos selecionados dos alunos, com vistas a manter uma documentação significativa do desenvolvimento de suas habilidades, tratando-se assim, do registro histórico e cumulativo da aprendizagem do aluno (Sistêlos, 1999), (Ramos, 2000).

Seguindo a linha de pensamento de Luckesi (Luckesi, 2000), como estratégia de avaliação, o portfólio, na fase de diagnóstico, permite a identificação mais significativa do conhecimento adquirido e das habilidades desenvolvidas pelo aluno, evidenciados através das diversas atividades realizadas durante todo o período escolar do aluno. Na fase decisória, as informações geradas permitem analisar a eficiência das estratégias didáticas aplicadas e do programa/currículo do curso; servem de subsídios para um diálogo entre professor, pais e aluno a respeito do progresso deste último; e de sugestões para as próximas atividades.

## 3.2. Acompanhamento do Aprendizado

O acompanhamento do aprendizado é essencial para que o processo avaliativo seja levado a termo com precisão e eficiência, fornecendo informações sobre o progresso e desenvolvimento do aluno.

Na área de educação, acompanhar o aprendizado significa determinar, através de instrumentos adequados, aspectos quantitativos e qualitativos do comportamento humano. Esses aspectos são variáveis da personalidade, tais como traços de caráter, de temperamento, capacidade de ajustamento, interesses, atitudes; ou aspectos relacionados

diretamente com a aprendizagem sistemática e aptidões, isto é, indicadores daquilo que o indivíduo já aprendeu ou está aprendendo.

O acompanhamento do aprendizado se dá segundo a medição de três espécies de características do aluno: aproveitamento, aptidões gerais e específicas e ajustamento pessoal-social (Linderman, 1986).

A medida de aproveitamento envolve a determinação do que o aluno aprendeu, isto é, quais são os conhecimentos que o aluno possui na ocasião em que é aplicado o teste. Com base em tais medidas, pode-se aferir a eficácia do programa de instrução e até que ponto o estudante alcançou o nível de desempenho dele esperado.

As medidas de aptidão permitem prever o que o aluno poderia aprender, dadas as experiências apropriadas de aprendizagem. Tais medidas indicam o nível de aprendizagem ou aproveitamento que se pode esperar do aluno em vista de suas aptidões gerais e específicas conhecidas.

Quando se mede o ajustamento pessoal e social, o objetivo de interesse é medir um certo número de características afetivas como honestidade, cooperatividade, lealdade e interesses. Como a função primordial da escola é desenvolver capacidades e aptidões cognitivas, o grande interesse pelas características afetivas justifica-se pela relação destas com o aproveitamento nas áreas acadêmicas.

Para realizar o acompanhamento segundo as características do aluno citadas acima, o professor tem uma série de instrumentos que numa ocasião ou noutra, se vê diretamente envolvido. São eles:

- **Testes Preparados pelo Professor:** são itens dissertativos e objetivos como provas, exercícios, tarefas, questionários, etc, que tem a função de medir a consecução de objetivos específicos de instrução no tocante a unidades particulares de trabalho. Seus resultados podem ser usados pelo professor para diagnosticar dificuldades de aprendizagem. Para ilustrar este uso dos resultados dos testes, considere a área de gramática inglesa. Os problemas de expressão oral e escrita podem às vezes relacionar-se com o fato de não terem sido compreendidas as funções de certas classes de palavras e suas relações com outras. Um teste de reconhecimento bem construído pode indicar, por exemplo, que, embora um determinado estudante não pareça ter problemas com o uso dos verbos, lhe é difícil identificar modificadores adverbiais. Portanto, uma testagem

diagnóstica deste gênero pode ajudar a melhorar a aprendizagem focalizando de maneira mais direta a origem da dificuldade e indicando o ponto em que se faz necessária uma educação corretiva;

- **Testes Padronizados pela Escola:** o objetivo aqui é avaliar o aproveitamento geral em uma ou mais áreas de disciplinas. O termo padronizado refere-se ao fato de que estes testes se destinam a serem aplicados, avaliados e interpretados dentro de condições-padrão. Devido a esta característica, eles são úteis para fazer comparações entre alunos ou grupos de alunos, mesmo quando os testes foram aplicados em diferentes ocasiões e lugares. Dessa forma, os resultados dos testes padronizados constituem uma base importante para a avaliação do progresso e aproveitamento dos alunos, auxiliando principalmente o professor no agrupamento de alunos, com vistas em obter uma instrução mais individualizada (Melchior, 1999);
- **Testes de Desempenho:** os testes de desempenho visam medir e acompanhar adequadamente a consecução de certos objetivos de instrução que não se podem medir através de testes elaborados pelo professor ou através de testes padronizados. Por exemplo, a capacidade de realizar operações em laboratórios de química, a velocidade e correção num curso de digitação, a execução adequada de uma aquarela em artes figurativas ou a de uma peça musical num instrumento;
- **Técnica da Observação:** É necessário que o professor registre as observações realizadas durante o processo de aprendizagem, para ter condições de ir redirecionando seu trabalho no sentido de ajudar os alunos a construir novos conhecimentos. Os registros de cada dia servem de subsídios para o professor ter em vista quais alunos não estão conseguindo avançar, quais necessitam retomar determinados aspectos do aprendizado, e enfim, quem são os que podem prosseguir com tarefas mais avançadas.

A observação é também importante como uma técnica adequada para a apreciação dos aspectos do desenvolvimento pessoal que não podem ser aferidos através de provas ou outros instrumentos avaliativos. Assim, o uso da observação pode favorecer o acompanhamento do aprendizado, principalmente por promover: a possibilidade de obter dados importantes que não podem ser constatados de outra forma;

facilidade de análise de todas as dimensões do comportamento humano; constatação da maneira como o aluno realiza as tarefas, etc.

- **Testes de Aptidão:** além de avaliar o que o aluno sabe ou o que ele aprendeu a fazer, é importante que em várias fases da sua experiência escolar se possa fazer alguma comparação entre o seu desempenho real e o seu potencial de êxito nas tarefas escolares de que vier a participar. Os testes de aptidões são utilizados para esta finalidade. Com os seus resultados, certas ações pedagógicas podem ser tomadas como por exemplo, colocar alunos em grupos apropriados de aprendizagem ou o planejamento de programas vocacionais mais eficientes.

Os resultados dos testes padronizados, testes de aptidão, testes elaborados pelo professor e registros de observações informais podem prestar um grande serviço para a tomada de decisão pedagógica, onde o professor pode averiguar o andamento do aprendizado dos alunos e revisar metas de ensino. Na realidade, segundo Linderman (Linderman, 1986), essa tomada de decisão pode representar:

- uma *operação preditiva*, ou seja, com base no desempenho presente e passado, deve-se formar um juízo sobre o possível sucesso ou fracasso de um estudante em várias atividades que ele empreenderá futuramente. Assim, todo professor, ao enfrentar uma nova turma de alunos no início do ano letivo, pode constatar que os resultados de testes e observações em cursos anteriores podem ser usados para fazer prognóstico sobre problemas que vários alunos terão probabilidade de enfrentar no decurso desse ano. Tais predições formam, portanto, uma base para um programa de ação preventiva a determinados alunos que terão probabilidade de apresentar problemas, de modo que tais problemas poderão ser combatidos com mais eficiência;
- uma *operação classificatória*, onde o professor classifica os alunos com base na consecução de certos objetivos escolares. Algumas classificações dizem respeito ao sucesso acadêmico dos alunos, enquanto outras focalizam os seus desempenhos no desenvolvimento de habilidades sociais, atitudes e características desejáveis de personalidade. Algumas classificações podem ser de natureza perfeitamente informal e feita, primariamente, para uso do próprio professor ao tomar nota dos progressos dos alunos; outras são mais formais e

podem ser comunicadas aos pais ou comissões/autoridades de ensino. Algumas dessas classificações, particularmente, as que se referem ao aproveitamento acadêmico, podem basear-se num considerável acervo de características objetivas; outras, pelo contrário, são inteiramente subjetivas e baseiam-se em impressões que o professor formou sobre o aluno graças ao contato presencial durante meses ou semanas. Algumas refletem, por parte do mestre, um sincero esforço para formar juízo acurado sobre as características e o progresso dos alunos; outras, são feitas de maneira bastante errôneas e não mostram a intenção de oferecer um retrato fiel do indivíduo.

As classificações usadas nas escolas para obter informações sobre características cognitivas e não cognitivas dos alunos estão susceptíveis a erros. Entre eles incluem-se os seguintes: a maioria dos professores, excessivamente atarefados, não tem tempo suficiente para realizar trabalhos cuidadosos para que as classificações tenham o máximo de validade; a pessoa que faz a classificação tende sempre a permitir que os seus sentimentos pessoais em relação ao aluno afetem a exatidão da classificação; alguns classificadores tendem a classificar quase todos os indivíduos numa posição mais ou menos mediana; as classificações relativas a diversos traços de um indivíduo tendem a ser todas altas ou todas baixas; algumas vezes, a característica a ser avaliada não se deixa observar facilmente pelo classificador; finalmente, uma fonte de infidedignidade das classificações é simplesmente a falta de conhecimento sobre a pessoa a ser classificada, em particular, no tocante aos traços específicos que é preciso avaliar.

Tais erros tendem a reduzir a validade das classificações.

### 3.2.1. Análise de Dados

Iniciemos esta seção com a seguinte afirmação de Johnson (Johnson, 2000): *“Educadores efetivos tomam decisões efetivas, decisões baseadas em informações precisas. Se conhecimento é poder, então conhecer as habilidades e atitudes dos estudantes significa dar poder aos educadores para julgar e acompanhar a evolução dos seus estudantes, promovendo as melhorias necessárias.”*

Quando os educadores estudam seus alunos, eles procuram respostas para questões como: como meus alunos estão se saindo nos testes padronizados? Quais são



suas dificuldades? Quais suas virtudes? Em que posso melhorar, para que meus alunos aprendam de forma efetiva? Mudar a estratégia de ensino? Aplicar mais testes? Dar mais oportunidades para a prática? Que características subjetivas meus alunos apresentam com mais intensidade? Como posso utilizar dados sobre avaliação e atividades, e até mesmo, dados como sexo, situação financeira, etnia, idade, e outros, para proporcionar um ensino mais direcionado?

Uma solução para estes questionamentos é colecionar dados dos estudantes e utilizá-los para tomar decisões mais elaboradas, onde o uso efetivo destes dados pode fornecer *feedback* necessário, elucidando os pontos fortes e fracos do processo educacional. Em especial, podemos destacar que o uso dos dados pode ajudar nos seguintes pontos:

- Medir o progresso dos alunos;
- Medir a efetividade dos programas;
- Avaliar a efetividade instrucional;
- Melhorar as estratégias de ensino;
- Guiar a alocação de recursos;
- Mostrar tendências;
- Construir perfis de alunos para melhor entender seus pontos fortes e fracos;
- E mais importante, construir oportunidades para acompanhar como os estudantes estão aprendendo e então usar esta informação para mudar sua estratégia de ensino (Boston, 2002).

Nesta direção, o processo de tomada de decisão baseado em dados (*Data-Driven Decision Making – D3M*) está se tornando uma tendência na educação. É crescente o número de professores, superintendentes e administradores escolares que estão utilizando este poderoso aliado, o dado, para favorecer os objetivos de ensino e da própria instituição de ensino no reconhecimento de suas metas e planos.

De acordo com Bernhardt (Bernhardt, 1999), em educação, D3M é o processo de coletar e analisar dados usando os resultados destas análises para promover melhorias na educação. Logo, é necessária a definição de quais dados serão importantes para o processo.

Na literatura, existe uma série de trabalhos que mostram quais dados são importantes e como guiar o processo de D3M. Exemplos podem ser vistos em (Keeney, 1998), (Adams, *et al.*, 2000), (Johnson, 2000), (Bernhardt, 2000), (Richardson, 2000) e (Protheroe, 2002).

Em se tratando do estudo dos dados, pode-se dividi-lo em duas partes: na primeira, abordaremos os tipos de dados; e na segunda, abordaremos a organização dos dados.

### **Tipos de Dados**

Existem basicamente dois tipos de dados, quantitativos e qualitativos. Cada tipo de dado prove um tipo de informação diferente, porém, ambos são vitais para o processo de D3M.

- **Quantitativos:** são dados como o resultado de testes padronizados, o número de faltas de um estudante, quantos estudantes estão cursando uma disciplina, qual o percentual de alunos acima da média, e outros;
- **Qualitativos:** incluem resultados de entrevistas, atividades extra-classe como trabalhos esportivos e participação em peças de teatro, informações sobre a situação financeira do aluno, região de onde vem, habilidades subjetivas como capacidade de liderança, criatividade e outras.

### **Organização dos Dados**

Os dados devem ser classificados nas seguintes categorias (Bernhardt, 2000):

- **Dados sobre Avaliação:** esta é a mais importante classe de dados que podem ser coletadas, pois são a partir dela que o professor tem a possibilidade de verificar em que pontos do aprendizado seus alunos precisam de reforço.

Estes dados devem ser coletados das múltiplas formas de avaliação utilizada, como testes padronizados, portfólios, exercícios, trabalhos, projetos, etc; e são essenciais para o desenvolvimento de um modelo de estudante baseado nas suas habilidades, conhecimentos, estilos de aprendizagem, realizações, motivações e outros;

- **Dados sobre o Estudante:** estes dados permitem que o professor saiba mais sobre o conjunto de alunos que ele possui, que não apenas seus resultados e habilidades. Tais dados incluem informações demográficas e comportamentais.

Informações demográficas diz respeito a características sócio-econômicas e geográficas dos alunos, tais como sexo, idade, etnia, status econômico, região de procedência, etc. Informações comportamentais são aquelas relacionadas à situação do estudante no curso. Em conjunto, estas informações são muito importantes pois elas descrevem a parte do sistema educacional sobre a qual se tem menos controle, clarificando questões como: a identificação de problemas sociais do estudante; aproveitamento por sexo, etnia e região derivando ações de ensino mais localizadas; identificação de perfis de estudantes desistentes com base nas informações do status econômico; identificação de perfis de alunos que mais trancam o curso com base em informações da região de procedência; e outros.

- **Dados Perceptivos:** alguns educadores em suas práticas, falham no sentido de não ouvirem as opiniões dos alunos sobre as práticas de ensino, metodologias, etc; e quais as suas perspectivas.

Os dados perceptivos podem ser agrupados de várias maneiras, tais como questionários, entrevistas e observações.

A reunião destas informações forma uma base sólida de dados para análises, permitindo a identificação de padrões que ajudam no desenvolvimento de estratégias de acompanhamento dos alunos. Em especial, pode-se entender a construção do conhecimento sendo criada pelos estudantes de forma mais abrangente que o tradicional e assim, promover mudanças mais efetivas no processo de ensino.

### 3.3. Acompanhamento do Aprendizado na Educação a Distância - EAD

Nesta seção, trataremos da problemática do acompanhamento do aprendizado em EAD. Inicialmente, apresentaremos o conceito de Educação a Distância. Na seqüência, mostraremos os desafios para se implementar esta modalidade de ensino. Seguindo, apresentaremos uma classificação das formas de acompanhamento em EAD e os trabalhos relacionados.

### 3.3.1. Educação a Distância

Em meados dos anos 60, procurou-se definir Educação a Distância – EAD pelo o que ela não era, tomando como referencial a educação presencial, onde eram estabelecidas comparações imediatas entre estas modalidades de ensino. Este comportamento não está completamente incorreto, mas promove um entendimento parcial do que é a EAD.

Estudos mais recentes apontam para uma conceituação mais precisa do que é a EAD. Passemos a observar com maior detalhes como os pesquisadores da área expressam o que consideram essencial para conceituar a EAD:

- G. Dohmem: *“Educação a Distância é uma forma sistematicamente organizada de autoestudo onde o aluno se instrui a partir do material de estudo que lhe é apresentado, onde o acompanhamento e a supervisão do sucesso do estudante são levados a cabo por um grupo de professores. Isto é possível de ser feito a distância, através da aplicação de meios de comunicação capazes de vencer longas distâncias. O oposto de “educação à distância” é a “educação direta” ou “educação face-a-face”: um tipo de educação que tem lugar com o contato direto entre professores e estudantes.”* (Nunes, 2000);
- G. Aretio: *“Educação a Distância é um sistema tecnológico de comunicação bidirecional (e massivo) utilizado como estratégia preferencial de ensino, substituindo a interação professor-aluno em sala de aula pela ação sistemática e conjunta de recursos didáticos e apoio de uma organização tutorial, propiciando a aprendizagem autônoma do estudante.”* (Neto, 1998);
- D. Bordenave: *“Educação a Distância é uma forma de organização de ensino-aprendizagem, na qual alunos de diversas idades e antecedentes estudem, quer em grupos, quer individualmente em seus lares, locais de trabalho ou outros lugares, com materiais auto-instrutivos distribuídos por meios de comunicação, garantida a possibilidade de comunicação com docentes, orientadores/tutores ou monitores.”* (Neto, 1998).

Tomando como base os conceitos acima apresentados, a EAD pode ser caracterizada por (Nascimento, 2002):

- Separação física entre professores e alunos;
- Auto-aprendizagem do aluno;
- Influência da organização educacional na preparação dos recursos didáticos;
- Utilização de meios de comunicação para unir professores e alunos e para transmissão de conteúdos e recursos didáticos;
- Possibilidade de atendimento a indivíduos geograficamente dispersos;
- O andamento do aprendizado é função do ritmo do aluno.

É importante observar que como a educação presencial, a educação a distância deve também possuir um compromisso pedagógico vinculado ao contexto histórico, político e cultural da sociedade (Azevedo, 2000). E as instituições que decidirem adotar esta modalidade de ensino, deverão reestruturar-se, alterando significativamente seus procedimentos. Conseqüentemente, não há como adotar a EAD sem alterar o papel do aluno e do professor. É preciso buscar novas formas de fazer educação desprendida dos velhos modelos, buscar novos paradigmas metodológicos, oferecer novos recursos e atividades pedagógicos, enfim, são necessárias muitas mudanças tanto na estrutura curricular, quanto nos critérios de planejamento, execução e avaliação dos programas de EAD.

Quanto às tecnologias aplicadas à educação a distância, atualmente são utilizadas todas as tecnologias disponíveis para transmissão de conhecimento, principalmente aquelas relacionadas à transmissão, armazenamento e apresentação das informações. Em especial, a Internet é destacadamente o meio mais importante que veio a viabilizar a proliferação de cursos à distância que temos nos dias atuais. Com ela, ao contrário do livro ou outros recursos, surgiram as comunidades virtuais e a “*inteligência coletiva*” definida por Lévy, onde passa a existir sinergia de competências, de modo que cada indivíduo pode dar sua contribuição, disponibilizar informações e trocar experiências em tempo real (Lévy, 1999).

A *World Wide Web* permite a publicação de informações de várias formas como textos, imagens, sons, animações, vídeos além do acesso a bancos de dados interativos. Adicionalmente, a Internet possibilita, como recursos de interação, tanto a comunicação

síncrona através do uso de *chats*<sup>1</sup> ou videoconferências, como a comunicação assíncrona utilizando as tecnologias de e-mail e grupos de discussões, e também a troca de informações e experiências entre indivíduos através de trocas de arquivos, sejam eles anexados aos correios eletrônicos, disponibilizados para *downloads*<sup>2</sup> em páginas HTML ou por uso de um protocolo de transferência de arquivos.

Todos estes recursos interativos providos pela Internet irão enriquecer o ambiente de aprendizagem permitindo o surgimento de ambientes cooperativos e colaborativos, ao contrário dos ambientes individualizados oferecidos pelas outras mídias empregadas na EAD como rádio, televisão, fitas de vídeo ou correspondências, onde a comunicação é unidirecional em que raramente ocorre a comunicação inversa. Todavia, para que realmente exista um ambiente colaborativo é necessário que todos os envolvidos estejam conscientes da necessidade de cooperar, particular ativamente e desenvolver o sentimento de comunidade (Nascimento, 2002).

### 3.3.2. Desafios da EAD

A educação a distância é uma modalidade de educação que apresenta desafios a serem enfrentados:

- Uma mudança cultural é exigida para que a educação a distância atinja seus objetivos educacionais uma vez que o aluno não está face-a-face ou presente no momento da aula. Logo, manter a motivação e a atenção é uma tarefa difícil dadas as outras atividades desempenhadas pelo estudante fora do ambiente de aprendizado e até mesmo pela presença de vários ambientes de entretenimento disponíveis;
- Para a elaboração de um curso, existe a necessidade de envolvimento de uma equipe multidisciplinar já que a intenção não é simplesmente transmitir um conteúdo que antes era transmitido numa sala de aula. O ideal é servir-se de todos os recursos disponíveis como imagens, animações, simuladores etc. Portanto, um professor ficaria responsável pelo planejamento do curso e elaboração dos conteúdos, um programador deverá desenvolver a aplicação *Web*

---

<sup>1</sup> O *chat* também é conhecido como bate-papo e corresponde a uma forma de comunicação onde cada indivíduo pode comunicar-se instantânea e textualmente com um conjunto de pessoas.

<sup>2</sup> *Download* é um meio de recuperar um arquivo disponibilizado num computador remoto.

e um *designer* gráfico será envolvido na apresentação e navegação de um curso a distância.

- Por fim, outro grande desafio a ser enfrentado é a avaliação e o acompanhamento do aprendizado do aluno devido à falta do contato presencial, onde a ausência da percepção do mestre quanto ao estado de compreensão de seus alunos pode levar ao insucesso de um curso a distância (Masetto, 2000).

A seguir, apresentaremos a problemática do acompanhamento do aprendizado em educação à distância.

### 3.3.3. Acompanhamento do Aprendizado em EAD

Hoje, quando se pensa em acompanhamento do aprendizado na educação a distância, percebe-se claramente que um longo caminho deve ser construído, e um dos maiores desafios é articular as inúmeras vantagens criadas pelas tecnologias modernas e recriar metodologias de avaliação e acompanhamento que efetivamente contenham alguma possibilidade de mudança qualitativa.

Um dos principais problemas sobre o acompanhamento em EAD é a falta de interações face a face entre professores e alunos. Dessa forma, em cursos à distância é necessário que as estruturas de acompanhamento sejam aperfeiçoadas e adaptadas, e que novos modelos ou estratégias que representem o status de aprendizagem do estudante remoto sejam elaborados. Neder (Vieira, 2002) destaca que o acompanhamento do aprendizado em EAD, embora possa se sustentar em princípios análogos ao da educação presencial, necessita de tratamento e considerações especiais. Esta argumentação da autora parte de algumas questões consideradas específicas da modalidade de EAD: a importância de desenvolver no estudante a autonomia crítica e desenvolver a capacidade de análise, tendo em vista a separação física entre o professor e o estudante. Souza (Souza, 1999) alerta que a prova, a forma mais comum de avaliação do aprendizado, não é suficiente para atestar os conhecimentos do aluno, pois se prende a um único resultado do processo de ensino-aprendizado. Por conseguinte, um professor na modalidade de EAD deve avaliar um aluno empregando os mecanismos tradicionais associados a outros meios, e assim propiciar uma avaliação realizada continuamente no

processo de ensino enquanto acompanha a construção e a evolução do aprendizado dos seus alunos. Guttierres e Prieto (Guttierres & Prieto, 1994) em suas reflexões sobre a avaliação na EAD, recomendam a proposição de uma avaliação alternativa orientada para a auto-avaliação do estudante. Segundo os autores, fazer uma avaliação alternativa significa fazer o diferente, a partir de propostas que visam a construção do conhecimento através da reelaboração, da experimentação e da aplicação, tendo em vista a reorientação, a correção e o estímulo ao estudante durante todo o processo.

Vale ressaltar que o processo de avaliação poderá ser realizado de maneira contínua e descritiva, possibilitando analisar como ocorre o desenvolvimento da aprendizagem do aluno, pela reflexão da sua experiência prática a partir dos referenciais teóricos apreendidos no curso.

Também se acredita que um sistema de acompanhamento, quando bem estruturado e com a utilização dos recursos tecnológicos para comunicação dos professores com os estudantes, poderá desempenhar um papel importante no processo de avaliação, principalmente no envio ou troca constante de *feedback*. Assim, as perguntas que surgem neste momento são: será que o atual sistema de acompanhamento é um meio eficiente que contribui para o processo de avaliação? De que forma? Será que por meio desse sistema os estudantes poderão, de alguma forma, superar suas dificuldades? E quanto aos *feedbacks*? São processos que contribuem efetivamente para que os professores possam acompanhar seus alunos? Como monitorar o aprendizado dos alunos num ambiente de EAD?

Neste sentido, alguns trabalhos foram propostos. Aqui, procuramos dividi-los em duas categorias: acompanhamento informal e acompanhamento alternativo.

### **Acompanhamento Informal**

Nesta categoria, são capturadas algumas características do aprendiz com o intuito de analisá-las de maneira análoga ao comportamento de um aluno de um curso presencial, onde a linguagem corporal, o grau de interesse, a participação e a sociabilidade podem ser vistas sob a forma da interação do aluno com o ambiente, provendo meios suficientes para avaliá-lo e fornecendo o *feedback* necessário sobre o acompanhamento do mesmo.



- Santos (Santos, 1999) enumera várias atitudes e comportamento de alunos em cursos a distância via Web que podem ser identificados, refletindo seus interesses e desempenhos:
  - Que caminhos foram percorridos sobre os conteúdos disponibilizados pelo professor?
  - Qual o grau de utilização e pesquisa de fontes suplementares fornecidos pelo professor?
  - Quais foram às contribuições e em que grau, na realização de tarefas cooperativas?
  - Com que freqüência eles contataram o professor por e-mail?
  - Contataram o professor somente próximo às avaliações?
  - Quais foram suas assiduidades e graus de participação em bate-papos, videoconferência, lista e fóruns de discussão?
  - Os trabalhos desenvolvidos pelos aprendizes mostraram boa utilização dos recursos educacionais disponíveis no curso?
  
- Menezes *et al.* (Menezes, *et al.*, 1999) acreditam que a observação do comportamento dos alunos durante o processo de interação com o ambiente pode captar o conhecimento acerca da aquisição de conhecimento dos mesmos permitindo uma complementação aos processos de avaliação existentes como provas, exercícios, etc; e destaca algumas atitudes passíveis de serem controladas em ambientes de aprendizado on-line:
  - Que caminhos alternativos foram percorridos em relação ao conteúdo didático proposto?
  - Quais fontes foram consultadas e com que freqüência?
  - O aluno utiliza fontes suplementares fornecidas pelo professor?
  - O aluno tem iniciativa de pesquisar fontes suplementares por conta própria?
  - O aluno interage com professores regularmente?
  - O ritmo de conjunto de alunos no consumo de material didático é regular?
  
- Alves *et al.* (Alves, *et al.*, 2002) destacam que outra atividade passível de monitoramento para critério de avaliação on-line, servindo também como

identificador de dificuldades, grau de interesse apresentado e nível de conhecimento obtido, é a relevância das mensagens postadas nos ambientes de EAD, pois podem fornecer informações valiosas sobre o desenvolvimento do curso e o grau de atividade mental do aluno;

- Rahkila e Karjalainen (Rahkila & Karjalainen, 1999) discutem a idéia de usar sistemas de *log* constituído pelo histórico do material didático acessado pelo estudante e assim podendo extrair informações de interesse do professor no que diz respeito ao tempo, ordem e freqüência de acesso dos estudantes aos materiais do curso.

### **Acompanhamento Alternativo**

Nesta categoria, as realizações feitas pelos alunos ao longo de um curso são criteriosamente avaliadas segundo uma série de itens mais subjetivos possibilitando *feedback* aos professores quanto às qualidades e dificuldades dos mesmos; além de que, os próprios alunos podem refletir e acompanhar seu próprio desempenho e verificar a aquisição de conhecimentos, habilidades e atitudes comprovadas através dos resultados de todos os seus trabalhos.

- Taminoto (Taminoto, 1998) se propôs a identificar alguns componentes básicos como registros para avaliação alternativa do aprendizado, acompanhando de forma sistemática e completa a evolução dos estudantes. Para isso, Taminoto definiu a seguinte taxonomia:
  - **Itens Históricos Educacionais:** representa as atividades e ações na qual o estudante está ou esteve envolvido. Isto inclui: *log* de atividades, descrição das atividades, *log* de comunicação, *quizes*, etc;
  - **Itens de Portfolio:** são os artefatos gerados pelos alunos que podem ser apresentados a outras pessoas como exemplo de suas habilidades ou como uma explanação de todo o seu *background* educacional;
  - **Itens de Avaliação:** inclui os resultados de projetos e atividades dos estudantes e as inferências sobre o nível de entendimento, motivação e progresso dos mesmos;
- Nascimento (Nascimento, 2002), baseado nos trabalhos de Sistêlos (Sistêlos, 1999), definiu uma avaliação autêntica baseado no método portfolio, onde todos

os artefatos produzidos pelo aluno, como trabalhos, estágios, exercícios, pesquisas, seminários, entre outros, são avaliados por itens subjetivos tais como clareza, completude, apropriação de conteúdo, capacidade de síntese, etc. Tais características refletem o aspecto de interesse e motivação do estudante diante das atividades postadas pelo professor, além de que o professor pode tomá-las como parâmetro para o acompanhamento do aprendizado de forma alternativa ao acompanhamento presencial como provas, testes padronizados e observação informal na sala de aula.

### 3.3.4. Ambientes Existentes

Existem diversos trabalhos que tratam do acompanhamento em EAD. Dentre as tecnologias mais utilizadas, podemos destacar o uso de sistemas tutores inteligentes<sup>3</sup>, agentes de softwares e banco de dados relacionais.

A seguir, apresentaremos alguns destes trabalhos.

- Menezes *et al.* (Menezes, *et al.*, 1999) propõem uma arquitetura multi-agentes para suportar a avaliação informal no ambiente AulaNet composta por três agentes: o Agente de Monitoração, que monitora as interações do aluno com o ambiente criando um histórico de navegações (*log* de ações) e percorrendo as listas de discussões a fim de verificar a participação dos alunos (*log* de interações); o Agente de Avaliação, que auxilia o professor na consulta e na interpretação do *log* de interações; e o Agente de Avaliação de Design que é capaz de indicar possíveis distorções no projeto instrucional;
- Souto *et al.* (Souto, *et al.*, 2001) propõem um sistema multi-agentes de coleta de *log* para o rastreamento das interações do aluno com o material instrucional seguido da identificação do padrão de comportamento cognitivo do aluno a partir da observação de suas interações com o ambiente;
- Jaques e Oliveira (Jaques & Oliveira, 1998) apresentam uma arquitetura multi-agente capaz de realizar o monitoramento das ferramentas de comunicação num

---

<sup>3</sup> Um sistema tutor busca apresentar um ensino individualizado a partir da operacionalização computacional das seguintes atividades: detecção do estado mental do estudante e estruturação de diferentes métodos instrucionais, onde a partir do estado detectado, possibilitar a escolha dos métodos instrucionais mais adequados, apresentado o conteúdo e a avaliação de maneira personalizada (Marietto & Omar, 2000).

ambiente de EAD, analisando as discussões que se encontram em andamento e disponibilizando ao professor as informações quantitativas sobre quais os assuntos mais discutidos, o grupo de alunos que mais dialogam entre si e outras informações que indicam o grau de participação dos alunos no curso;

- Alves *et al.* (Alves, *et al.*, 2002) desenvolveram uma estratégia de avaliação onde a partir de valores padrões armazenados em banco de dados para os aspectos monitorados pelo sistema como número e relevância das mensagens postadas em fóruns de discussão, bate-papo e correio eletrônico, e o tempo de conexão com o ambiente, o sistema é capaz de representar os aspectos comportamentais do aluno e modelar o estado de aprendizado do mesmo;
- Silva (Silva, 2002) desenvolveu um método de acompanhamento do estudante baseado em técnicas de *Data Mining* para o ambiente DE-MAW (Seno & Vieira, 2002), onde as interações do aluno com o ambiente e a realização e os resultados das avaliações foram armazenados num Data Warehouse<sup>4</sup> de acompanhamento. A partir dos dados neste repositório, o professor pode extrair padrões relativos ao comportamento dos alunos utilizando tarefas de mineração de dados;
- Nascimento (Nascimento, 2002) propõe um sistema tutor inteligente acoplado a um portfolio eletrônico baseado em banco de dados temporal e ativo para promover uma avaliação personalizada (de acordo com o estado cognitivo do aprendiz) e autêntica, onde, dada sua natureza pedagógica do portfolio, permite o reconhecimento mais significativo das habilidades desenvolvidas pelo aluno e possibilita uma análise mais detalhada sobre a eficiência das estratégias didáticas aplicadas e do programa do curso;
- Silva e Fernandes (Silva & Fernandes, 2000) propõem uma sociedade de agentes para promover o acompanhamento do aprendizado em EAD onde se destaca o agente Aprendiz, cuja tarefa é auxiliar o aluno em suas atividades; e o agente Facilitador, cuja tarefa é assistir o professor na confecção de instrumentos de avaliação e atividades de acompanhamento como por exemplo, corrigir a parte não subjetiva das avaliações, adequar o conteúdo aos perfis de alunos, ajustar os

---

<sup>4</sup> Um Data Warehouse é um repositório centralizado de dados desenvolvido especialmente para atender a consultas gerenciais.

instrumentos às habilidades e comportamentos diagnosticados, auxiliar na detecção de aspectos subjetivos dos estudantes, dentre outras;

- Otsuka *et al.* (Otsuka, *et al.*, 2002) apresentam uma ambiente para apoiar a aprendizagem baseada na resolução de problemas. Dessa forma, foram desenvolvidas diversas ferramentas integradas para apoiar o desenvolvimento de atividades. Dentre elas, podemos destacar a ferramenta Portfolio, que é a área onde um aprendiz ou grupo de aprendizes pode organizar suas informações, a fim de comunicar ao grupo e/ou ao professor a disponibilização de artefatos e os resultado de seus trabalhos e receber comentários e sugestões de forma colaborativa e construtiva, motivando o aprendiz a entrar num ciclo de revisões e possibilitando a oportunidade de construir e depurar os novos conhecimentos;
- Rodrigues e Geyer (Rodrigues & Geyer, 2001) apresentam um sistema multi-agente de avaliação que integra a avaliação formal e a avaliação informal, procurando identificar problemas de aprendizagem ocorridos durante a interação do aluno no ambiente de ensino a distância e durante a realização de trabalhos, reconhecendo quando ele está precisando de ajuda ou então quando a estratégia de ensino não corresponde ao seu perfil. Depois de reconhecido o problema de aprendizagem, o agente de avaliação deve solucioná-lo, apontando uma nova estratégia de ensino;
- Guedes *et al.* (Guedes, *et al.*, 2001) propõem uma ferramenta de análise textual que utiliza técnicas de *Text Mining*<sup>5</sup> para auxiliar o professor a avaliar textos construídos em ambientes de ensino-aprendizagem. Os textos são obtidos por meio de discussões virtuais a respeito de temas determinados através de um *chat*. A ferramenta auxilia o professor a determinar se os aprendizes estão realmente discutindo o tema em questão, se não estão perdendo tempo com trivialidades e se estão aprendendo e produzindo conhecimento, realizando algumas tarefas com esta finalidade, como identificação de alunos e suas interações, e apresentação e classificação das palavras e frases utilizadas, disponibilizando os resultados ao professor para análises;

---

<sup>5</sup> *Text mining* é uma área derivada da mineração de dados cujo intuito é encontrar padrões em dados não estruturados, gerando novos conhecimentos. Alguns de seus objetivos são extrair informações chaves de textos, organizar documentos por assunto, encontrar temas predominantes em conjuntos de documentos e pesquisar documentos através de consultas (Gotthard & Seiffert, 1997).

- Machado e Becker (Machado & Becker, 2002) propõem um modelo de descoberta de padrões de navegação em um ambiente de educação a distância cuja finalidade é verificar e avaliar a realização do processo de aprendizagem de acordo com a ordem e seqüência dos acessos realizados durante a navegação no ambiente de ensino e de acordo com o padrão de comportamento de navegação que reflete a maneira com que o aluno executou suas atividades. Assim, através destas descobertas, pode-se oferecer recursos que permitam a comparação e avaliação da estrutura atual do *site* educativo e seu uso real.

### 3.3.5. Análise dos Resultados

Como visto, a tarefa de acompanhar o aprendizado em educação à distância é bastante complexa. Porém, cabem algumas considerações.

Os trabalhos de Menezes *et al.* (Menezes, *et al.*, 1999), Souto *et al.* (Souto, *et al.*, 2001), Jaques e Oliveira (Jaques & Oliveira, 1998) e Alves *et al.* (Alves, *et al.*, 2002) se limitam a acompanhar as características informais dos alunos, como as interações dos alunos com o ambiente de EAD e as interações dos alunos com o material didático instrucional. Estes trabalhos não contemplam os aspectos alternativos da avaliação, onde características mais subjetivas poderiam ser analisadas e confrontadas com os resultados das avaliações.

Já os trabalhos de Nascimento (Nascimento, 2002), Otsuka *et al.* (Otsuka, *et al.*, 2002) e Silva e Fernandes (Silva & Fernandes, 2000) se preocupam com os fatores alternativos do acompanhamento, porém, deixam a desejar no aspecto informal do acompanhamento.

Pouco se tem feito em termos de sistemas de apoio a decisão. Na maioria dos trabalhos aqui apresentados, a tarefa de acompanhamento é feita de forma padronizada, sem que o professor possa decidir quais aspectos lhe interessam ou quais informações deseja confrontar. Apenas o trabalho de Silva (Silva, 2002) se destina a esta temática. Porém, apresenta a limitação de não contemplar aspectos alternativos da avaliação.

O trabalho de Rodrigues e Geyer (Rodrigues & Geyer, 2001) tem a preocupação de um acompanhamento mais elaborado, onde os aspectos formais, como trabalhos, provas, exercícios, e outros; e os aspectos informais, como as interações do aluno com o

ambiente de EAD, são tratados e analisados. Porém, não existe análise de características subjetivas dos alunos.

Nenhum trabalho apresentado mostrou-se vinculado ao processo de D3M. Tal processo auxilia o professor nas suas atividades de acompanhamento e diagnóstico de problemas de aprendizagem, não somente relacionadas aos resultados das avaliações, mas também relacionadas a outros fatores como dados demográficos, comportamentais e perceptivos.

Podemos notar, portanto, que diante desta análise, não existe um conjunto de padrão de requisitos para acompanhar o aprendizado. Além disso, todos os trabalhos apresentados contemplam apenas um subconjunto de características dos alunos, deixando de lado outras características que também poderiam ser analisadas para favorecer um acompanhamento mais abrangente da evolução cognitiva e afetiva dos alunos.

A Tabela 3.1 abaixo sumariza a análise apresentada. O resultado obtido pode ser interpretado da seguinte forma: “X” indica a existência do recurso, e “-“ indica a inexistência do recurso.

TRABALHO	FATORES DE ACOMPANHAMENTO				
	Alternativo	Informal	Demográfico	Comportamental	Avaliação
Menezes <i>et al.</i>	-	X	-	-	-
Souto <i>et al.</i>	-	X	-	-	-
Jacques & Oliveira	-	X	-	-	-
Alves <i>et al.</i>	-	X	-	-	-
Silva	-	X	X	-	X
Nascimento	X	-	-	-	X
Silva & Fernandes	X	-	-	-	X
Otsuka <i>et al.</i>	X	-	-	-	X
Rodrigues & Geyer	-	X	-	-	X
Guedes <i>et al.</i>	-	X	-	X	-
Machado & Becker	-	X	X	-	-

**Tabela 3.1.** Análise Comparativa dos Resultados

### 3.4. Considerações Finais

Neste capítulo, apresentamos o papel da avaliação no processo de ensino e aprendizagem. Mostramos também alguns fatores que podem ser considerados no acompanhamento do aprendiz e a problemática do acompanhamento do aprendiz na modalidade de educação à distância.

Toda a base teórica apresentada aqui e no capítulo anterior nos servirá para propormos no capítulo seguinte, um modelo de acompanhamento de aprendiz baseado em técnicas de mineração de dados auxiliando o professor no processo de tomada de decisão de natureza pedagógica.



### *UMA ESTRATÉGIA PARA APLICAR MINERAÇÃO DE DADOS NO ACOMPANHAMENTO DO APRENDIZADO EM EAD*

Neste capítulo, é apresentada uma estratégia para aplicar técnicas de Mineração de Dados, quando se deseja obter informações sobre o acompanhamento do aprendizado em ambiente de EAD. Sua exposição está dividida da seguinte forma: inicialmente, são feitas algumas considerações; em seguida, mostraremos os requisitos de uma estratégia de acompanhamento baseado em Mineração de Dados; na seqüência, apresentaremos as etapas que comporão a estratégia.

#### 4.1. Considerações Iniciais

Existem muitos sistemas computacionais voltados para a realização da EAD. Tais sistemas abordam grande parte do processo ensino/aprendizagem: a elaboração e disponibilização do material didático, a apresentação das aulas, a comunicação entre os participantes do curso e a avaliação do aprendizado.

Porém, avaliar e acompanhar o aprendizado dos alunos em um curso a distancia não são tarefas triviais, pois envolvem teorias pedagógicas, nas quais os professores estruturam seus cursos, questões tecnológicas, como rastreamento e autenticação de alunos e apoio à tomada de decisão por parte do professor, mediante situações problemáticas na dinâmica de ensino/aprendizagem (Silva, 2002).

O ensino visa uma evolução do modo de pensar e do comportamento do aluno com relação a um determinado tema, e espera-se que o aluno demonstre essa evolução em suas atitudes durante o curso. A detecção desta evolução, que se pode tomar como o acompanhamento do aluno, envolve uma análise da postura do aluno diante da metodologia de ensino, visando medir:

- **O grau de interesse do aluno:** numa sala de aula tradicional, o contato face-a-face pode ser a principal forma de perceber o interesse do aluno, embora num

ambiente de EAD também se possa medi-lo pela percepção do quanto o aluno consulta materiais que abordam o tema e pelas suas perguntas e comentários;

- **O grau de satisfação do aluno com o material didático:** as reações do aluno ao material usado para ensinar podem demonstrar o grau de adequação do mesmo ao aprendiz. Por exemplo, em cursos a distância, se o aluno consulta mais os *links* externos ao curso do que as aulas do próprio curso, pode ser que o material didático não esteja satisfazendo o aprendiz. Como reflexo dessa medida, o professor pode reformular seu material didático para as próximas turmas;
- **O grau de compreensão e desenvolvimento de habilidades do aluno:** o professor utiliza instrumentos de avaliação com o intuito de medir o progresso dos alunos. Ele pode usar os resultados e verificar quais são as deficiências que devem ser suprimidas e não simplesmente julgá-lo pelo seu resultado;
- **O grau de interação do aluno com os demais alunos e com o professor:** o relacionamento entre os atores do processo educacional pode indicar falhas na dinâmica de passagem, recepção e fixação do conhecimento. O aluno que não faz perguntas pode estar sendo coibido socialmente pelas atitudes de seus colegas e/ou professor. E o professor que não promove o diálogo pode estar perdendo o *feedback* sobre o curso.

A diferença entre o acompanhamento na educação presencial e na educação a distância está na observação realizada pelo professor, que não pode mais contar com o “corpo a corpo” da sala de aula. O uso do computador torna possível capturar algumas características do comportamento do aprendiz a distância, e analisá-las de uma maneira análoga ao comportamento de um aluno de um curso presencial. O grau de interesse, a participação, o comportamento social, podem ser vistos pela ótica computacional, considerando, basicamente, as interações do aluno com o ambiente de ensino. Por exemplo, o grau de interesse e a participação do aluno podem ser refletidos pela frequência com que o mesmo acessa o curso, enquanto que o comportamento social pode ser delineado pelo rastreamento de suas comunicações.

A proposta deste trabalho é a de aplicar uma estratégia baseada em técnicas de *Data Mining* para favorecer o acompanhamento do aprendizado, onde uma coleção de

dados sobre o aluno pode ser analisada e padrões comportamentais ou características interessantes dos alunos podem ser descobertas.

O uso de *Data Mining* é justificado segundo vários motivos:

- À medida que um curso à distância é concebido para atender ao público remoto, o contingente de alunos pode crescer consideravelmente, e, portanto, a tarefa de acompanhar o aprendiz pode se tornar árdua;
- Normalmente os cursos de EAD mantêm seus dados em bancos de dados e a natureza histórica destes dados poderia ser útil para análises prospectivas;
- Pouco se tem feito em termos de sistemas de suporte a decisão em EAD. A maioria dos ambientes existentes não oferece recursos para apoiar decisões mais elaboradas;
- As escolas estão descobrindo o poder dos dados na tomada de decisão e promover melhorias. Seguem alguns relatos:
  - Segundo Kinder (Kinder, 2000), os educadores estão aprendendo a usar os dados provenientes de diversas fontes como currículos, práticas educacionais, avaliação, entre outros, para auxiliar no diagnóstico de problemas ou determinar soluções eficazes.
  - Johnson (Johnson, 2000) destacou que a implementação de um programa de coleta e análise de dados pode levar a melhorias na educação como nenhuma outra inovação educacional o fez no último século;
  - Schmoker (Sparks *apud* Schmoker, 2000) destaca o papel da análise de dados para conduzir os objetivos de uma instituição de forma coerente e indicar os pontos de melhorias;
  - Bernhardt (Bernhardt, 2001) acredita que as decisões baseadas em dados históricos ajudam os educadores a enxergar quem são seus alunos e que qualidades e dificuldades eles compartilham. Com isso, os professores podem examinar suas práticas mais efetivamente;
  - Armstrong & Anthes (Armstrong & Anthes, 2001) dizem que as escolas e professores podem usar os dados de estudantes para medir o progresso e o impacto dos programas e métodos e tomar decisões sobre o currículo, o ensino e conseqüentemente a aprendizagem.

## 4.2. Requisitos da Estratégia

Uma estratégia de acompanhamento do aprendizado em EAD baseado em técnicas de Mineração de Dados deve considerar uma série de requisitos fundamentais para que sua utilização seja aplicada.

A seguir, são descritos os requisitos básicos levantados, juntamente com uma breve explicação de cada um.

### **R1 – A estratégia deve contemplar as etapas de KDD**

As etapas de seleção, pré-processamento, transformação, garimpagem e análise e assimilação dos resultados, apresentadas na seção 2.2, são igualmente importantes para qualquer processo de descoberta de conhecimento e devem estar presentes na estratégia.

### **R2 – A estratégia deve atender as primitivas de *Data Mining***

A estratégia deve ser interativa com o minerador. Para isso, deve atender as primitivas de *Data Mining* que foram apresentadas na seção 2.3.3 e incluem:

- **dados relevantes à tarefa:** o usuário pode selecionar parte do banco de dados a ser utilizado para análises;
- **tipo de conhecimento minerado:** o usuário tem a possibilidade de selecionar que tarefa de mineração ele deseja utilizar, e assim gerar um modelo de conhecimento específico para cada tarefa;
- **medida de interesse:** o usuário pode informar as medidas capazes de filtrar as regras produzidas ou para avaliar o conhecimento minerado;
- **avaliação e visualização dos resultados:** o usuário pode visualizar os padrões extraídos e assim averiguar se tais padrões podem ser utilizados em sua vida prática.

### **R3 – O usuário utilizará técnicas de Mineração de Dados para buscar padrões nos dados**

O acompanhamento do aprendizado deverá ser apoiado com a utilização de uma ferramenta de Mineração de Dados que irá buscar conhecimento relevante nos dados. Com isso, o usuário poderá usar os resultados da mineração para acionar algumas

medidas como por exemplo, aplicar uma nova estratégia de ensino ou alterar a seleção de futuros alunos para o curso.

#### **R4 – A estratégia deve dispor dados para o acompanhamento informal**

A estratégia deve dispor dados sobre as interações do aluno com o ambiente de EAD e com os materiais didáticos disponibilizados pelo professor, provendo dados para análises de características como graus de interesse, de participação e de sociabilidade.

Aqui, podemos destacar duas categorias básicas de informações informais (Vrasidas & McIsaac, 1999):

- **interações aluno-aluno e aluno-professor:** são aquelas em que o aluno interage com outros integrantes do curso, nos dando informações sobre a participação do aluno, motivação e grau de sociabilidade;
- **interações aluno-material:** são aquelas em que o aluno interage com o material didático. Este tipo de informação possibilita ao professor de detectar padrões de acesso e aceitação/rejeição de materiais, etc;

#### **R5 – A estratégia deve dispor dados para o acompanhamento alternativo**

A estratégia deverá dispor dados sobre os itens de julgamento das características subjetivas dos alunos proporcionadas pela avaliação autêntica. Dessa forma, o professor terá condições de analisar outras habilidades que não apenas àquelas relacionadas aos aspectos técnicos, mas também pessoais e de condicionamento afetivo.

Podem ser considerados itens de avaliação como: capacidade de síntese, capacidade de crítica, criatividade, e outros que se julgam necessários.

#### **R6 – A estratégia deverá dispor dados sobre avaliações, dados demográficos e dados comportamentais para análises prospectivas**

A estratégia deverá dispor de dados sobre as avaliações dos alunos como provas, exercícios, trabalhos, etc; dados sobre o próprio aluno como sexo, idade, local onde mora, *status* financeiro, etc; e dados sobre comportamentais como a sua situação no curso, número de reprovações, número de suspensões e outros. A união destas informações juntamente com os aspectos informais e alternativos do acompanhamento, proverá um *background* completo das realizações, habilidades e características dos

alunos. E toda esta coleção de dados, em grandes quantidades, poderá gerar padrões e tendências interessantes que favoreçam o acompanhamento do aprendizado.

#### **R7 – A estratégia deve possibilitar o cruzamento dos dados coletados**

Todos os dados do acompanhamento informal, alternativo e os dados sobre avaliações, demográficos e comportamentais poderão ser cruzadas sem nenhuma restrição, e assim servir como entrada para que algum algoritmo de mineração de dados gere conhecimento acerca das habilidades, atitudes e comportamentos dos alunos, permitindo o professor acompanhar do aprendizado dos mesmos. Ou seja, o usuário poderá utilizar tal conhecimento para diagnosticar as influências de um determinado fator sobre um outro. Dessa forma, ele poderá tomar as atitudes cabíveis para que o aprendizado seja melhorado.

### 4.3. Descrição da Estratégia

A nossa estratégia para apoiar decisões pedagógicas no acompanhamento do aprendizado está dividida em duas fases: a **fase de planejamento** e a **fase de análise de dados**. A seguir, descreveremos cada fase.

#### 4.3.1. Fase de Planejamento

Na fase de planejamento, devemos analisar os seguintes aspectos: que permissões terão os usuários para usar um sistema de apoio à tomada de decisão pedagógica para o acompanhamento do aprendizado em EAD; quais condições para definição e modelagem de dados de interesse o sistema de EAD dispõe; e que tipo de mineração deve ser realizada.

Sob o ponto de vista de usuários do sistema, deve-se delimitar o escopo de utilização dos mesmos, já que, como em qualquer outro sistema, alguns usuários têm permissões para manipular apenas uma parte dos dados que lhes são de interesse e responsabilidade. Logo, a definição de quais dados podem ser disponibilizadas para quais usuários deve ser considerada. Por exemplo, um professor somente poderá

minerar dados de acompanhamento relativo as suas turmas. Já um coordenador poderá analisar dados do seu curso; e um reitor poderia minerar dados sobre a instituição como um todo em busca do perfil de aluno que possui.

Sob o ponto de vista da definição e da modelagem dos dados de interesse, possíveis situações do ambiente inicial de projeto devem ser analisada, a fim de que o planejamento seja efetuado com o aproveitamento total dos recursos existentes. Partindo deste princípio, uma das seguintes situações ocorre ao ser iniciado um projeto visando o acompanhamento do aprendizado através do uso de Mineração de Dados:

### **1. Existe um ambiente com base de dados que dá suporte ao uso de mineração de dados**

Neste caso, existe uma base de dados, e ela já está preparada para a utilização da mineração de dados. Portanto, a etapa de preparação de dados presente no requisito R1 já é satisfeita.

Um exemplo de um mecanismo utilizado nesta situação pode ser encontrado em Silva (Silva, 2002), onde existe um *Data Warehouse* de acompanhamento e um ambiente de extração de conhecimento em cubos mineráveis<sup>6</sup>. O usuário pode formular consultas sobre os dados armazenados neste *Data Warehouse*, criando cubos de dados. Na seqüência, o usuário tem a possibilidade de minerar o cubo criado e gerar padrões.

Nesta situação ainda é relevante conhecer os aspectos relacionados à disponibilidade de dados para os usuários. Se não existir controle de acesso e restrição de dados, deve-se implementar tal mecanismo. Para tal, o uso de SQL poderá ajudar, através do comando *grant*, ou então, restringindo o acesso pela interface do usuário.

### **2- Existe um ambiente com base de dados que não dá suporte ao uso de mineração de dados**

Neste caso, existe uma base de dados, porém a mesma não foi concebida para suportar o uso de mineração de dados. Para tal situação, deve-se seguir os seguintes passos:

- 1º - Defina padrões de consultas para os grupos de usuários do sistema de apoio a decisão. Padrões de consultas são as possibilidades de consultas que cada

---

<sup>6</sup> Cubo minerável é aquele criado para ser utilizado numa tarefa de mineração de dados qualquer.

grupo de usuário poderá efetuar ao utilizar o sistema. Por exemplo, professores poderão consultar dados sobre suas disciplinas e turmas. Por sua vez, coordenadores somente consultarão os seus cursos;

- **2º** - Defina um conjunto de visões para atender aos padrões de consultas. Uma visão pode ser vista como uma relação contendo algum dado que contemple ou o requisito R4 (dado do acompanhamento informal) ou o R5 (dado do acompanhamento alternativo) ou o R6 (dado demográfico ou comportamental ou sobre avaliação); porém nunca deve contemplar mais de um destes requisitos. O motivo para esta restrição é simples: como os requisitos R4, R5 e R6 contemplam dados de acompanhamento distintos, o usuário tem a possibilidade de correlacionar as informações que desejar. Se uma visão contemplar, por exemplo, o requisito R4 e o R5, o usuário terá que correlacionar dados de acompanhamento informal e alternativo respectivamente. E isto pode não ser desejado ou pode não fazer sentido.

Destacamos dois mecanismos para a implementação do conjunto de visões:

- **Abordagem virtual**, com a utilização de *Views* dos SGBD's. Com ele, é possível criar tabelas virtuais acessíveis apenas por usuários autorizados. Para cada padrão de consulta, pode-se criar um conjunto de *Views* (um para cada fator de acompanhamento presente ou no requisito R4 ou no R5 ou no R6). Além do mais, com o uso de SQL, a tarefa de preparação de dados presente no requisito R1 poderá ser satisfeita: a tarefa de limpeza pode ser obtida com o uso do comando *select*; as tarefas de redução e agregação podem ser obtidas com o uso das funções de agregação; as tarefas de discretização e transformação podem ser resolvidas com o uso de *Stored Procedures*; e assim sucessivamente;
- **Abordagem materializada**, com a utilização de um sistema de *Data Warehousing*. Tais sistemas contêm ferramentas de extração, limpeza, agregação e consolidação dos dados de interesse num *Data Warehouse* de acompanhamento. Desta forma, a etapa de preparação de dados especificada no requisito R1 seria satisfeita.



### 3- Inexiste base de dados

Nesta situação, o sistema de EAD torna disponível apenas as aulas e não possui nenhum tipo de dado armazenado em bancos de dados. Neste caso, deve ser desenvolvida uma base de dados que atenda aos requisitos da estratégia, de tal maneira que dê suporte ao uso de mineração de dados e contemple todas as informações passíveis de análises já discutidas.

Logicamente, como estamos tratando do acompanhamento do aprendizado, estamos definindo a granularidade mínima de análise como sendo as realizações, características e atitudes pertencentes ao aluno. Além do mais, a identificação do aluno deve ser camuflada como forma de segurança e ética.

Ainda sob o ponto de vista de definição de informações de interesse, nas situações 1 e 2 acima, podem ocorrer os seguintes casos:

**Caso 1:** Quando o requisito R4 não é satisfeito, isto indica que no sistema de EAD não foi previsto o acompanhamento informal. Para remediar tal situação, pode-se adotar as seguintes posições:

- **1º** - Utilize os *logs* dos servidores *web* para conhecer os acessos aos materiais didáticos disponíveis pelo professor através de sua *url*. No *log* do servidor *web* é possível descobrir informações como usuário, *url*, data de acesso e duração do acesso. Assim, é possível capturar dados sobre as interações do aluno com os recursos didáticos;
- **2º** - Crie meios de capturar as interações do aluno com o ambiente de EAD, como participações em *chats*, listas de discussões ou correio eletrônico.

Um mecanismo para a implementação do acompanhamento informal é a utilização de tabelas de dados (*Table* dos SGBD's). Com elas, as interações do aluno com o ambiente podem ser convenientemente armazenadas.

Para a situação 1, seria suficiente incrementar a base de dados já preparada para a mineração de dados com estas tabelas; para a situação 2, pode-se utilizar tais tabelas na criação de *views* ou cubos para os padrões de consultas.

**Caso 2:** Quando o requisito R5 não é satisfeito, o acompanhamento alternativo não foi previsto. Ou seja, no sistema de EAD, a única forma de avaliação prevista é baseada em notas. Para contornar tal situação, é necessária uma mudança estrutural na instituição de EAD. Quanto se resolve adotar a avaliação autêntica, todos os envolvidos no processo educacional devem estar dispostos a acreditar nas vantagens que este método de avaliação pode trazer. Resolvidos tais conflitos, a instituição deve pré-definir os artefatos que poderão ser produzidos pelos alunos, bem como os itens de julgamento que avaliarão tais artefatos. Padronizados os artefatos e os itens de avaliação, o passo seguinte é disponibilizá-lo para o sistema de apoio a decisão. Uma forma de se implementar tal tarefa é também com a utilização de tabelas dos SGBD's. Com elas, os portfolios dos alunos contendo todas as suas realizações - artefatos, bem como a avaliação dos mesmos, poderiam estar armazenadas apropriadamente.

Mais uma vez, para a situação 1, seria suficiente incrementar a base de dados já preparada para a mineração de dados com estas tabelas; para a situação 2, pode-se utilizar tais tabelas na criação de *views* ou cubos para os padrões de consultas.

**Caso 3:** Quando o requisito R6 não é satisfeito, quer dizer que os dados a serem utilizados no processo de D3M convencional não foram disponibilizados. Para remediar tal situação, deve-se adotar as seguintes estratégias:

- **1º** - Considerando que não existam dados demográficos e comportamentais:
  - Se existe um cadastro de alunos, incremente o cadastro com os dados demográficos e comportamentais de interesse, caso o cadastro não possua tais dados;
  - Se não existe um cadastro de alunos, crie um cadastro de alunos com dados demográficos e comportamentais de interesse. Para isso, pode-se usar uma tabela (*Table* dos SGBD's);
  - Disponibilize os dados demográficos utilizando o mecanismo de visões citado anteriormente, de acordo com o padrão de consulta do usuário.
- **2º** - Considerando que não existem dados sobre avaliações, a estratégia é criar meios de armazenar as tarefas que expressam o desempenho do aluno em termos de atividades propostas pelo professor. Cada atividade proposta pelo professor pode ter um resultado. Por exemplo, a participação ou não em uma

reunião, a nota de um trabalho, etc. Considerando isso, esse tipo de informação depende das atividades escolhidas pelo professor e da forma que ele escolheu para validá-las, ou seja, o critério para decidir se elas foram cumpridas ou não pelos alunos.

A Tabela 4.1 apresenta alguns exemplos de atividades e suas formas de validação.

<u>Atividade</u>	<u>Validação</u>
Prova	Nota
Exercícios	Nota
Projetos/Trabalhos	Nota
Leitura	Entrega de resumo
Debates	Nº de mensagens enviadas
Seminários	Nota para material produzido e apresentação
Conferência	Nº de mensagens enviadas

**Tabela 4.1.** Atividades Propostas e suas Validações

Para atualização do modelo de dados aqui proposto, considerando situação 1, se o banco de dados preparado para a mineração não previa dados relacionadas ao D3M, incremente-o com tais dados. Já na situação 2, pode-se usar as tabelas e criar visões ou cubos para atender aos padrões de consultas definidos para os usuários.

Como resultado desta fase de planejamento, temos um **MAD - Modelo Analítico de Dados**, refletindo um completo *background* escolar dos estudantes. Assim, diferentemente dos modelos apresentado na seção 3.3.4 do capítulo 3, um conjunto de dados demográficos, comportamentais, informais, alternativos e dados sobre avaliação está disponível num único ambiente. Adicionalmente, com a formalização de um MAD, o usuário terá a possibilidade de navegar e selecionar os dados de interesse que estão disponíveis para uma consulta em questão. Sendo assim, o usuário usará o MAD para escolher os atributos que serão utilizados para análises prospectivas.

Fechando a fase de planejamento, o último tópico a ser analisado é a disponibilidade de tarefas de mineração a serem utilizadas pelos usuários. Para esta questão, pode-se utilizar qualquer tipo técnica. Porém, o interessante é se fazer um estudo das necessidades dos usuários e das potencialidades que cada tarefa de

Mineração de Dados pode lhes trazer, avaliando assim, se uma determinada tarefa pode ser utilizada ou não. Por exemplo, a mineração por agrupamento poderia ser utilizada para agrupar um conjunto de alunos segundo alguma característica; a mineração por associação poderia ser utilizada para verificar associações entre características distintas (informal *versus* alternativo; alternativa *versus* demográfica; etc) dos estudantes; a tarefa de classificação poderia ser usada para dispor perfis de alunos segundo algumas habilidades; e assim sucessivamente.

No que diz respeito aos mecanismos de implementação, existe hoje uma gama de softwares de mineração de dados. Todos eles proporcionam uma série de algoritmos já implementados e diversas ferramentas de visualização de resultados. Dentre elas, podemos destacar:

- DBMiner (DBMiner, 2002): utilizado para descoberta de conhecimento em bancos de dados relacionais, onde a integração com o SGBD é feita através da linguagem DMQL (*Data Mining Query Language*). A ferramenta permite o acesso a um BD via ODBC (*Open Database Connectivity*) e pode-se também utiliza-lo integrado com ferramentas OLAP. A grande vantagem do DBMiner está em suas excelentes interfaces gráficas para visualização dos resultados;
- IBM Intelligent Miner (IBM, 2002): desenvolvido pela IBM e acoplado ao SGBD DB2, implementa diversos algoritmos de DM para regras de associação, classificação, agrupamento, entre outros. Além disso, o grande ponto forte da ferramenta é as funções de preparação de dados e suas interfaces de visualização de resultados. Estas funções permitem a totalização de valores, a eliminação de atributos desconhecidos, discretização, etc;
- Framework WEKA (Witten & Frank, 2000): O framework *WEKA (Waikato Environment for Knowledge Analysis)* agrega uma variedade de algoritmos de indução de conhecimento. Porém seu grande ponto forte é o fato de o mesmo ser *freeware*, além de sua condição de *framework* permitir o fácil desenvolvimento de aplicações utilizando seus componentes.

Para finalizar, a fase de planejamento é ilustrada no diagrama de atividades da figura 4.1 a seguir.

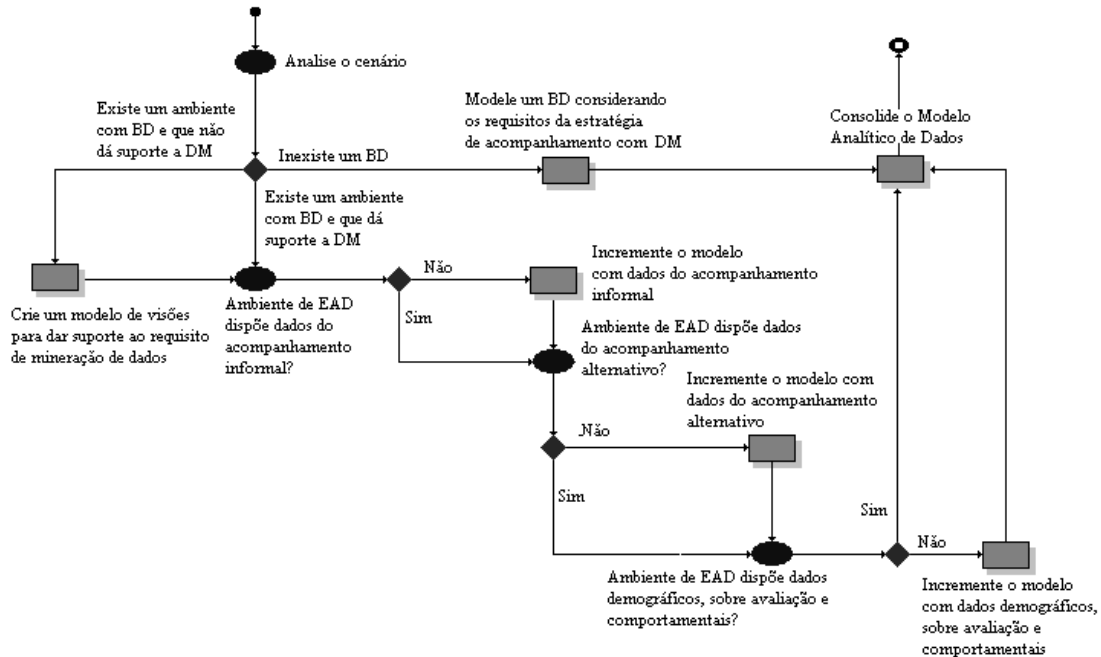


Figura 4.1. Diagrama de Atividades da Fase de Planejamento

### 4.3.2. Fase de Análise de Dados

Propomos nesta fase uma abordagem semi-automática para extração de conhecimento baseado em técnicas de *Data Mining* com o intuito de gerar regras, padrões, tendências e/ou associações relativos às características do estudante armazenadas historicamente no banco de dados e disponíveis através do MAD. Tais padrões podem revelar um conhecimento útil no que diz respeito ao acompanhamento do aprendizado, apoiando os educadores no processo de tomada de decisão e gerando novas oportunidades estratégicas ao providenciar as seguintes capacidades:

- Identificação das causas de problemas, não apenas dos sintomas;
- Possibilidade do cruzamento de informações demográficas, comportamentais, informais, alternativas e sobre avaliação propiciando uma análise mais elaborada que o tradicional;
- Troca de pressentimentos e hipóteses por fatos;
- Avaliação de necessidades e alocação de recursos para satisfazê-las;
- Verificação do cumprimento das metas de ensino;
- Identificação de perfis de potenciais alunos desistentes;

- Determinação de atividades mais eficazes no curso;
- Questões que tradicionalmente requeriam análises extensivas podem agora ser respondidas diretamente através dos dados.

Achamos conveniente dividir esta fase de Análise de Dados em três etapas: *Etapa de Coleta de Dados, Etapa de Mineração de Dados e Etapa de Análise e Incorporação dos Resultados.*

### **Etapa de Coleta de Dados**

Na Etapa de Coleta de Dados, o usuário irá utilizar as informações previstas num MAD, e formular uma consulta com os dados de seu interesse, de acordo com algum padrão de consulta disponível para o mesmo. Por exemplo, suponhamos que um MAD tenha sido formado com as seguintes informações:

- ***Acompanhamento Informal:***
  - **Interação via chats:** número de participações de um aluno numa seção de *chat*;
  - **Dedicação:** registra os horários de cada acesso do aluno ao ambiente de EAD;
  - **Interação com o material didático:** número de acessos dos alunos por material didático;
- ***Acompanhamento Alternativo:***
  - **Aproveitamento por Item de Avaliação:** registra o desempenho por itens de julgamento;
- ***Dados Demográficos:***
  - **Dados pessoais:** cadastro de dados pessoais do aluno como sexo, idade, região de procedência, atividade profissional e outras;
- ***Dados Comportamentais:***
  - **Reprovações:** indica o número de reprovações que o aluno obteve durante o curso;
  - **Situação:** indica se o aluno está efetivo no curso; ou se trancou o curso; ou se abandonou o curso;
- ***Dados sobre Avaliações:***
  - **Aproveitamento em atividades:** contém dados sobre o desempenho do aluno nas atividades.

A Tabela 4.2 sumariza as informações disponíveis neste MAD.

<u>Classe da Informação</u>	<u>Informação Disponível</u>	<u>Atributo</u>
Acompanhamento informal	Interação através de <i>chat</i>	Nº de participações
Acompanhamento informal	Dedicação	Nº de horas/dia
Acompanhamento informal	Interação com material didático	Nº de acessos/material
Acompanhamento alternativo	Aproveitamento por item de avaliação	Desempenho
Dados demográficos	Dados pessoais	Sexo, idade e renda
Dados comportamentais	Reprovações	Nº de reprovações/disciplina
Dados comportamentais	Situação	Efetivo/trancado/abandono
Dados sobre avaliação	Aproveitamento em atividades	Desempenho/atividade

**Tabela 4.2.** Informações Disponíveis no MAD-Exemplo

Um professor poderia verificar a relação existente entre o sexo e idade do aluno com o número de reprovações do mesmo em sua disciplina. Ou talvez desejasse correlacionar todos os aspectos do acompanhamento informal com os dados sobre a avaliação tipo “prova” dos estudantes. Para isso, ele selecionaria as visões/atributos de seu interesse e informaria alguns parâmetros para a consulta em questão.

Como meios de implementação, as estratégias variam de acordo com a ferramenta de mineração de dados selecionada para uso:

- **1º** - Se a ferramenta de DM selecionada for do tipo fortemente acoplada, disponibilizando mecanismos de elaboração de consultas já integrada com a base de dados e com as ferramentas de *Data Mining*, seria interessante aproveitar tais recursos e disponibilizar o seu sistema de apoio à tomada de decisão utilizando estas características;
- **2º** - Se a ferramenta de DM não disponibilizar mecanismos de elaboração de consultas, o seu sistema de DM deve prover meios de realizar a seleção. Portanto, esta é uma solução fracamente acoplada, ou seja, a ferramenta apenas implementa os algoritmos de mineração e utiliza dados a partir de arquivos *flat files* para a geração de regras. Nesta situação, como mecanismo de implementação, é interessante utilizar comandos SQL dinâmicos, onde as cláusulas do comando *select* seriam montadas dinamicamente através da seleção de visões/atributos e parâmetros num MAD. As condições da cláusula *where* seriam especificadas de acordo com o padrão de consulta do usuário acrescido

das possíveis junções entre as tabelas/visões criadas para o MAD. Além do mais, no próprio SQL dinâmico, poderiam ser especificados alguns pré-processamentos de dados, como por exemplo, agregações obtidas pelo uso das funções de agregação e a eliminação de dados faltantes obtida pelo próprio *select*. Com o posterior conjunto-resposta – *resultset* – mais alguns pré-processamentos poderiam ser realizados, como por exemplo, a discretização de valores contínuos ou reduções. Estes dados já preparados poderiam ser então transformados num arquivo *flat* obedecendo a sintaxe definida pelo fabricante da ferramenta de DM;

- 3º - Se a ferramenta de DM for do tipo sem acoplamento, ela não se aplica, já que para a nossa estratégia, sempre estamos utilizando recursos de sistemas de banco de dados, como *view*, *table*, *data warehouse* e *stored procedures*.

Sendo assim, a primitiva de Mineração de Dados “dados relevantes a tarefa” especificada no requisito R2 é atendido. Ou seja, o usuário trabalhará apenas com uma parte do banco de dados que lhe interessar em uma dada situação.

### **Etapa de Mineração de Dados**

Na Etapa de Mineração de Dados, as tarefas de mineração previstas na fase de planejamento devem ser colocadas à disposição do usuário. Além do mais, as primitivas de *Data Mining* “tipo de conhecimento a ser minerado” e “medida de interesse” descritas no requisito R2 devem ser atendidas, bem como a etapa de garimpagem presente no requisito R1 e o requisito R3. Ou seja, o usuário, depois de realizar uma consulta num MAD a fim de selecionar os dados de interesse, tem a possibilidade de escolher que padrão de conhecimento deseja saber e que tipos de regras devem ser descobertas, de acordo com as medidas de interesse disponíveis, como suporte, confiança, etc.

A aplicação de tarefas de mineração de dados sobre os dados armazenados ao longo do tempo, relativos ao acompanhamento do aluno, poderá encontrar padrões de comportamento e habilidades que estejam ligados ao processo de aprendizagem. E esses padrões podem ser úteis no sentido de ajudar o professor a compreender melhor seus alunos.



## Etapa de Análise e Incorporação dos Resultados

Esta é a etapa onde o usuário tem a possibilidade de visualizar o conhecimento minerado e decidir se tal conhecimento é útil ou não para o seu trabalho de acompanhamento. Portanto, esta etapa contempla a primitiva “avaliação e visualização dos resultados” descrita no requisito R2.

A Figura 4.2 apresenta um exemplo de análise sobre o resultado de uma mineração. Os círculos indicam atributos selecionados para a mineração, as setas indicam padrões encontrados e os retângulos são conseqüências da ação sobre as regras geradas pela descoberta de conhecimento.

Tal resultado poderia ser acionado pelo professor, ou até mesmo por um agente inteligente de software ou um sistema tutor inteligente<sup>7</sup>. Tais sistemas poderiam fazer uso dos resultados da mineração para melhorar o acompanhamento de um aluno ou grupo de alunos. Ou seja, o conhecimento produzido poderia facilitar a decisão sobre qual estratégia de ensino adotar para melhorar a performance dos alunos ou detectar problemas.

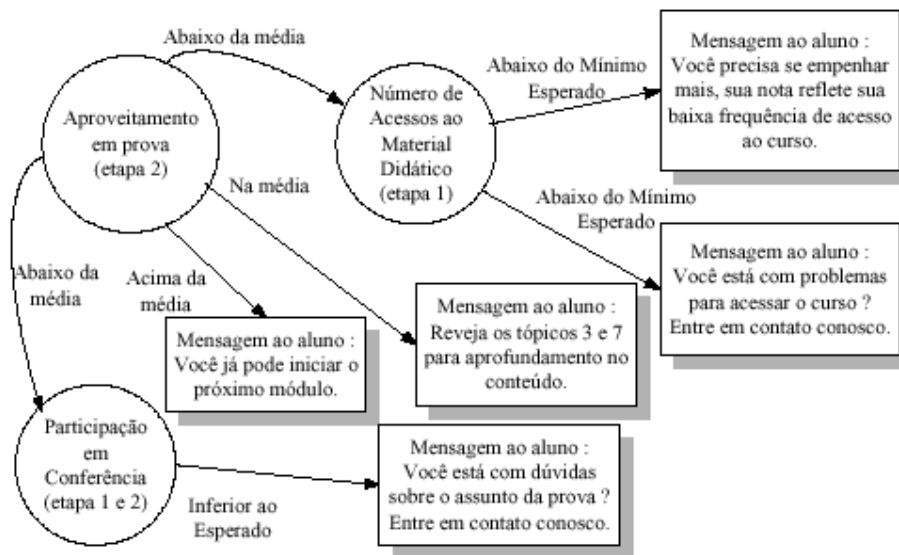


Figura 4.2. Exemplo de uma Análise de Dados (Silva, 2002)

A seguir, a Figura 4.3 sumariza as atividades realizadas na fase de Análise de Dados.

<sup>7</sup> Um sistema tutor inteligente é um software que instrui pela apresentação de conceitos, que vão se tornando mais complexo de acordo como andamento do aluno (Nascimento, 2002).

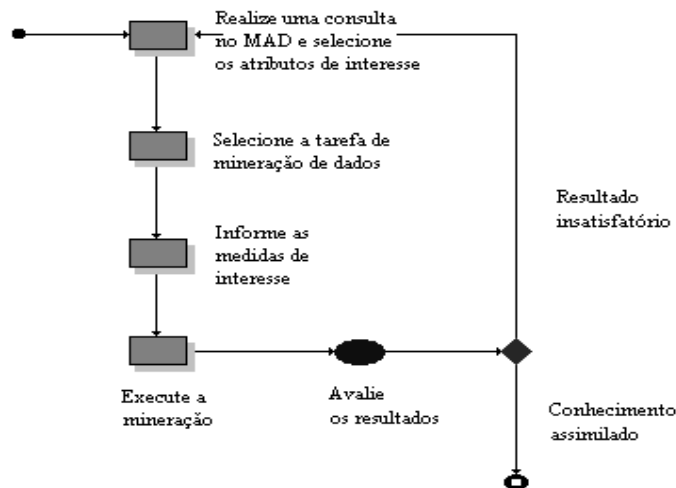


Figura 4.3. Diagrama de Atividades da Fase de Análise de Dados

## 4.4. Considerações Finais

Neste capítulo, apresentamos uma estratégia para apoiar decisões pedagógicas no acompanhamento do aprendizado em EAD. O grande diferencial desta estratégia em comparação com os trabalhos relacionados ao tema - acompanhamento do aprendizado em EAD - mostrados na seção 3.3.4 do capítulo 3, é que a mesma prevê as informações passíveis de serem analisadas em um único repositório chamado de Modelo Analítico de Dados. Além disso, permite a utilização de tecnologias de *Data Mining* para gerar conhecimento novo e útil favorecendo as decisões relacionadas ao processo de ensino/aprendizagem.

### *MIDAS-POETA: UM SISTEMA DE APOIO À TOMADA DE DECISÃO PARA O AMBIENTE PORTFOLIO-TUTOR*

Este capítulo apresenta o sistema Midas-Poeta. Tal sistema é concebido utilizando-se a estratégia de apoio à decisão pedagógica para o acompanhamento do aprendizado mostrado no capítulo anterior. Inicialmente, fazemos uma rápida explanação do sistema Portfolio-Tutor, ambiente utilizado para o desenvolvimento do sistema de apoio a decisão. Depois, apresentaremos os aspectos de análise, projeto e implementação do Midas-Poeta.

#### 5.1. O Sistema Portfolio-Tutor

O sistema Portfolio-Tutor (Nascimento, 2002) foi resultado de pesquisas desenvolvidas na Universidade Federal de Alagoas. O objetivo deste ambiente de EAD foi o de proporcionar um sistema tutor (camada Tutor-ACVA) acoplado a um portfolio eletrônico de dados (camada Portfolio Eletrônico).

O Portfolio-Tutor faz parte do projeto ACVA - Arquitetura de uma Classe<sup>8</sup> Virtual Adaptativa (Hernandez-Dominguez, 1998), cujo propósito é a construção de um ambiente de ensino/aprendizagem com enfoque na criação de um ambiente colaborativo de aprendizagem e no ensino personalizado.

##### **A Camada Portfolio Eletrônico**

A camada Portfolio Eletrônico tem como principal objetivo auxiliar o professor nas funcionalidades de registro e acompanhamento de atividades de alunos. Sua idéia foi fundamentada pela proposta pedagógica do POETA (Portfolio Eletrônico Temporal e Ativo) (Sistêlos, 1999), onde o portfolio do aluno é subdividido em portfolio de

---

<sup>8</sup> Usa-se o termo Classe de forma equivalente ao termo Aula.

trabalho e portfolio de apresentação. No portfolio de trabalho são registradas todas as atividades atribuídas pelo professor aos seus alunos, por unidades da disciplina, e classificadas segundo os artefatos utilizados pela instituição de ensino. Os documentos gerados por cada aluno são arquivados e avaliados pelo professor conforme os itens de avaliação subjetivos e pré-estabelecidos. Os melhores trabalhos, selecionados em conjunto pelo aluno e professor, são disponibilizados no portfolio de apresentação para acesso do público em geral.

Ao utilizar o Portfolio-Tutor, a instituição educacional deve definir os tipos de evidências (artefatos) que podem ser utilizadas pelos professores, bem como os critérios específicos que devem ser empregados na avaliação de cada uma.

Exemplos de possíveis artefatos dos alunos seriam:

- Projetos de pesquisa;
- Críticas de artigos;
- Experiências de laboratório;
- Visitas de campo;
- Debates;
- Seminários;
- Apresentação de peças teatrais;
- Resumos, etc.

Considerando os itens que poderiam ser empregados na avaliação destes artefatos, têm-se:

- Apropriação de conteúdo;
- Clareza de exposição;
- Capacidade crítica;
- Capacidade de sintetização;
- Criatividade;
- Vocabulário diversificado;
- Participação ativa em debates;
- Motivação;
- Capacidade de liderança, etc.

## **A Camada Tutor-ACVA**

A camada Tutor-ACVA tem como principal objetivo auxiliar o professor no processo de transmissão de conhecimento, portanto, é a responsável por tratar as sessões de ensino propriamente ditas.

A camada Tutor-ACVA fornece um ensino personalizado, evitando a homogeneidade com que os alunos são tratados em sala de aula tradicional (Hernandez-Dominguez & Silva, 2000). Assim, o Tutor considera uma classe virtual formada por grupos heterogêneos de estudantes. Seguindo esta proposta, o perfil do aluno é subdividido em *perfil de grupo* e *zona de comportamento*. O perfil de grupo corresponde ao nível de conhecimento apresentado pelo aluno – Básico, Intermediário ou Avançado, em relação aos conteúdos apresentados. Enquanto que, a depender do desempenho obtido durante a sessão de ensino, as zonas de comportamento consideradas são: Superior, Normal e Crítica.

Os conteúdos ensinados em cada perfil de grupo e zona de comportamento são definidos segundo uma base de domínio composta de três níveis: domínio, módulo conceitual e recursos didáticos (Cabral & Giraffa, 2000), (Kawasaki, 2000) e (Resende & Barros, 2000). O domínio contém a hierarquia de conhecimento existente. O módulo conceitual contém o conjunto de conceitos de um conhecimento existente. E os recursos didáticos são o conjunto de recursos a serem utilizados na aprendizagem do aluno.

A fim de proporcionar um histórico das habilidades do aluno adquirido ao longo do tempo, o Tutor acrescenta ao modelo do aluno todos os conceitos já aprendidos e o histórico de navegação que registra todos os recursos didáticos utilizados pelo mesmo.

A mobilidade do aluno também é gerenciada pelo Tutor e se dá de modo intergrupar – mudança entre grupos, e intragrupal – mudança de zona de comportamento. Inicialmente, é feito um pré-teste para verificar o nível de conhecimento do aluno. A depender do resultado deste pré-teste, o aluno é inserido em um dos grupos (Básico, Intermediário ou Avançado) e numa zona de comportamento (Superior, Normal ou Crítica). A mobilidade se dá entre níveis de um perfil de grupo. Depois que o aluno superar o nível Avançado de uma zona, ele está pronto para mudar de grupo. A Figura 5.1 abaixo ilustra esta situação.

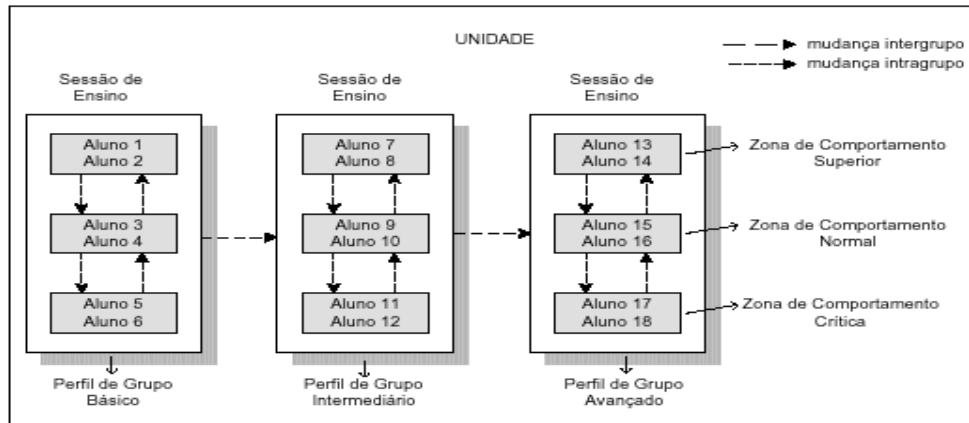


Figura 5.1. Distribuição dos Alunos de uma Aula Virtual Adaptativa (Nascimento, 2002)

### A Base de Dados do Sistema Portfolio-Tutor

A base de dados do sistema Portfolio-Tutor foi implementada no gerenciador de banco de dados *Microsoft SQL Server 7.0*. A fim de proporcionar uma visão mais modular, dividimos o esquema lógico do banco de dados em quatro partes: *Usuário*, *Estrutura*, *Portfolio* e *Tutor*. *Usuário* representa como os usuários do sistema foram modelados. A parte *Estrutura* diz respeito ao sistema de matrícula e divisão de disciplina em unidades. *Portfolio* e *Tutor* representam a modelagem do portfolio eletrônico e do tutor, respectivamente.

A Figura 5.2 ilustra a parte *Usuário* do banco de dados.

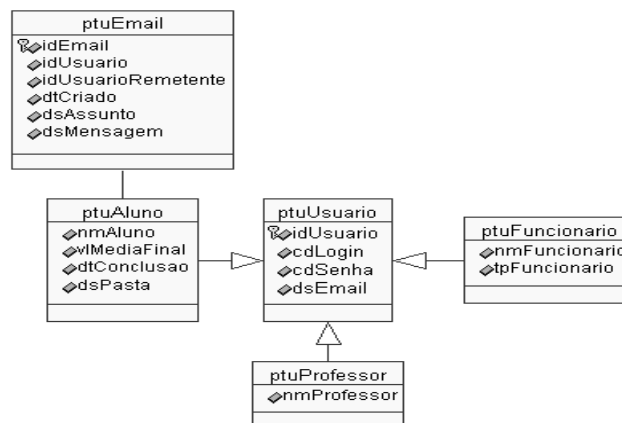


Figura 5.2. Banco de Dados do Portfolio-Tutor (Parte Usuário)

A tabela *ptuUsuario* mantém informações de *login* do usuário; as tabelas *ptuProfessor*, *ptuFuncionario* e *ptuAluno* são especializações de *ptuUsuario* e

representam os professores, funcionários e alunos cadastrados no sistema. A tabela *ptuEmail* registra as interações por correio eletrônico entre os alunos da instituição de ensino.

A Figura 5.3 ilustra a porção *Estrutura* do banco de dados: as tabelas *ptuCurso* e *ptuCursoDisciplina* registram, respectivamente, os cursos da instituição e as disciplinas de cada curso. A tabela *ptuDiscTurma* contém as turmas, por disciplina, de um dado professor. Já a tabela *ptuMatriculaDiscTurma* armazena as matrículas em disciplina dos alunos da instituição.

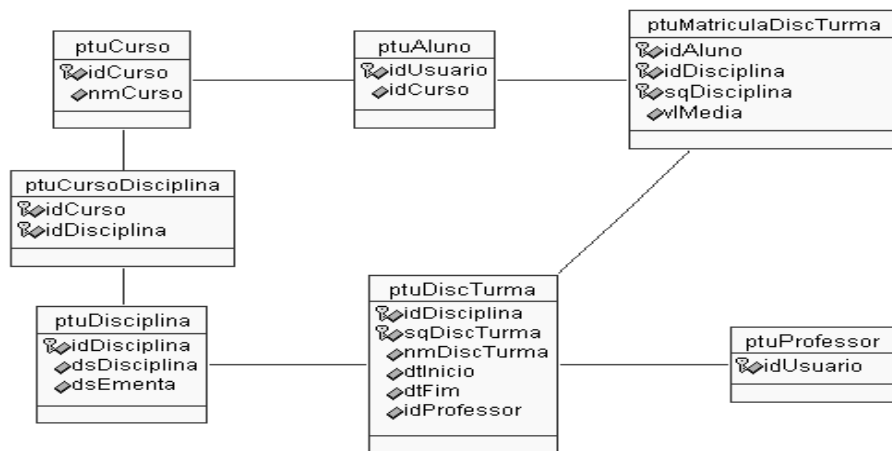


Figura 5.3. Banco de Dados do Portfolio-Tutor (Parte Estrutura)

A Figura 5.4 abaixo apresenta a parte de Portfolio: as tabelas *ptuUnidade*, *ptuAtividade* e *ptuAtividadeAluno* são as unidades e atividades planejadas em cada disciplina, e as atividades realizadas pelo aluno em cada unidade, respectivamente. Para compor o portfolio do aluno, cada atividade refere um artefato (*ptuArtefato*), que é avaliado segundo os itens de avaliação padronizados pela escola (*ptuItemAvaliação*). Para um artefato, porém, podem existir diversas atividades associadas. O professor julga as atividades atribuindo notas aos itens de avaliação (*ptuAvaliacaoItemAvaliacao*). Para cada atividade realizada pelo aluno, são entregues os documentos (*ptuDocumento*) que comprovam a sua realização. São exemplos de documentos: as respostas de um exercício, um resumo, um desenho, uma prova, enfim, qualquer artefato que represente a habilidade que o aluno adquiriu. Em cada unidade planejada na disciplina, é sumarizada a média da unidade (*ptuDesempenhoUnidade*). Isto permite verificar se o aluno está apto naquela unidade.

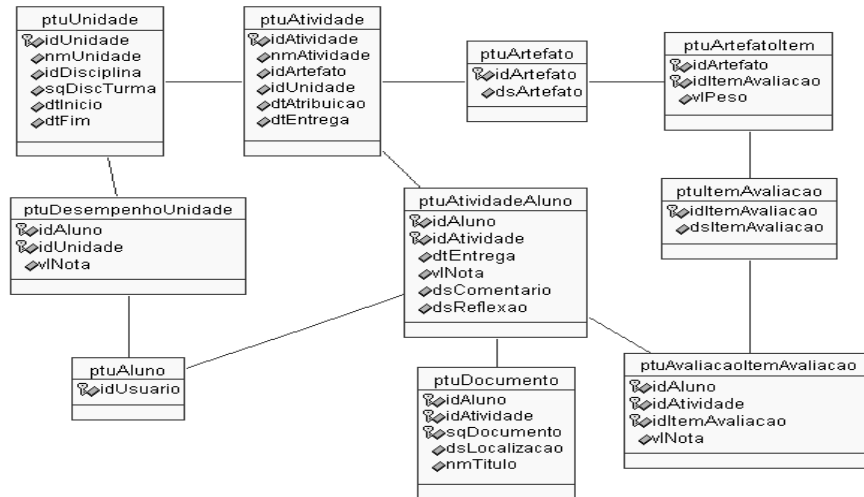


Figura 5.4. Banco de Dados do Portfolio-Tutor (Parte Portfolio)

A Figura 5.5 representa a parte *Tutor*.

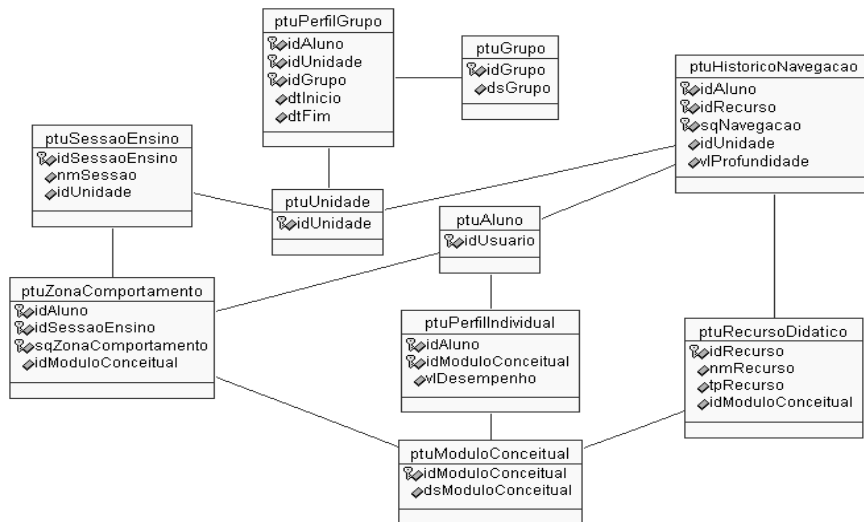


Figura 5.5. Banco de Dados do Portfolio-Tutor (Tutor)

Segundo o esquema lógico da Figura 5.5, a unidade é dividida em sessões de ensino (*ptuSessaoEnsino*) que atende a uma determinada zona de comportamento (*ptuZonaComportamento*). Cada sessão de ensino é constituída de módulos conceituais (*ptuModuloConceitual*). Para cada módulo conceitual, os conceitos são apresentados por recursos didáticos (*ptuRecursoDidatico*). O aluno, por sua vez, participa de uma zona de comportamento e de um perfil de grupo (*ptuPerfilGrupo*), e apresenta um perfil individual (*ptuPerfilIndividual*), que se constitui de todos os módulos conceituais



aprendidos, e do histórico de navegação (*ptuHistoricoNavegacao*) dos recursos didáticos já utilizado pelo mesmo.

### **Limitação do Sistema Portfolio-Tutor**

As informações capturadas e armazenadas neste sistema podem ser úteis para responder a questões do tipo:

- Em que ordem os alunos acessam os tópicos ou páginas de conteúdos?
- Os alunos seguem o caminho padrão para acessar o material?
- O grau de interesse do aluno pelo curso/disciplina sofreu alteração?
- Quais alunos interagem mais entre si?
- Quais assuntos foram mais debatidos e por quais alunos?
- Qual o desempenho do aluno na disciplina?
- O desempenho do aluno foi satisfatório? Em quais itens de avaliação?

Pode-se responder estas questões utilizando consultas em linguagem SQL. Porém, podem existir outras perguntas interessantes para o professor que seriam muito complicadas de serem formuladas em SQL, pois exigiriam análises mais abrangentes, considerando, muitas vezes, o comportamento de dados ao longo do tempo. Por exemplo, verificar a influência no aprendizado de fatores como a comunicação (através do uso de e-mails), a motivação (pelo acesso ao material didático) e o estudo (pelo resultado das avaliações dos artefatos). O maior obstáculo para responder a essas perguntas através de consultas em SQL é encontrar parâmetros para esta consulta.

Adicionalmente, com a utilização do portfolio, é gerada uma enorme quantidade de dados ao longo do tempo. Ao analisar estes dados convenientemente, pode-se extrair informações úteis (padrões, tendências, associações, regras, etc) que favoreçam o acompanhamento do aprendizado dos alunos. E o sistema Portfolio-Tutor não dá suporte a este tipo de análise, apresentando a limitação de não contemplar um Sistema de Apoio à Decisão - SAD, que auxilie o professor, de modo mais elaborado, em suas ações estratégicas de acompanhamento.

Com um SAD acessando uma base de dados preparada especialmente para suportar tais análises, o professor poderia ser amparado com informações efetivas e verificar se as práticas de ensino estão sendo conduzidas adequadamente.

Descreveremos nas seções seguintes o sistema Midas-Poeta, um sistema de apoio à tomada de decisão pedagógica no acompanhamento do aprendizado para o Portfolio-Tutor. Tal sistema deve fazer uso da estratégia apresentada no capítulo anterior e, portanto, tem como pontos fortes:

- A definição de um conjunto completo de dados de acompanhamento (um Modelo Analítico de Dados - MAD);
- A flexibilidade que o professor tem para escolher e confrontar os dados de acompanhamento (já que ele pode selecionar os atributos do MAD que lhe interessar) e analisar a influência de alguns atributos sobre outros;
- A utilização de técnicas de mineração de dados para descobrir padrões de conhecimento embutidos nos dados.

A Figura 5.6 ilustra uma arquitetura de alto nível do sistema Midas-Poeta e sua relação como Portfolio-Tutor.

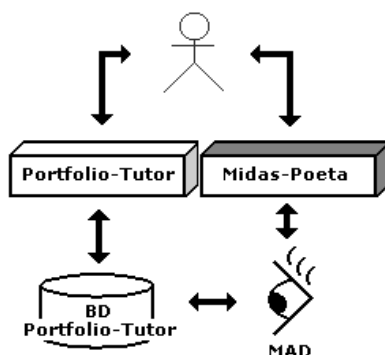


Figura 5.6. Arquitetura de Alto Nível do Midas-Poeta

O sistema Midas-Poeta é uma camada de software acoplada ao Portfolio-Tutor fazendo uso de sua base de dados para preparar o MAD. O MAD é constituído por um conjunto de visões sobre esta base de dados. Assim, o usuário do Midas-Poeta terá acesso ao MAD para realizar as suas consultas.

## 5.2. O Sistema Midas-Poeta

A seguir, detalharemos os aspectos de análise, projeto e implementação do sistema Midas-Poeta.

## 5.2.1. Análise do Midas-Poeta

Nesta seção, apresentaremos os requisitos funcionais<sup>9</sup> e não funcionais<sup>10</sup>, identificaremos os usuários do sistema e apresentaremos o diagrama de *use-cases*.

### 5.2.1.1. Requisitos Funcionais

**RF1 – O sistema deve dispor um conjunto de visões que contemple as variáveis de acompanhamento discutidas no Capítulo 4.**

O objetivo do sistema aqui proposto é o de apoiar as decisões pedagógicas no acompanhamento do aprendizado. Para isso, deve existir um Modelo Analítico de Dados – MAD que agregue todas as variáveis passíveis de análise. Cada variável deve ser mapeada numa visão e deve ser coletada a partir da base de dados do Portfolio-Tutor.

**RF2 – O usuário deve realizar uma consulta baseada no MAD.**

O usuário utilizará o Modelo Analítico de Dados para definir os atributos de interesse de uma consulta em questão. Em seguida, o sistema deverá construir a consulta do usuário dinamicamente em linguagem SQL.

**RF3 – O sistema deverá dispor tarefas de mineração de dados para descobrir padrões nos dados gerados a partir da consulta do usuário ao MAD.**

Realizada a seleção dos atributos de interesse no MAD, o sistema deverá permitir que o usuário minere o resultado de sua consulta. Para isso, devem estar disponíveis as tarefas de mineração por associação, classificação e agrupamento.

**RF4 – O resultado da mineração deverá ser exibido para avaliações posteriores.**

---

<sup>9</sup> Os requisitos funcionais denotam as funcionalidades a serem construídas pela equipe de desenvolvimento.

<sup>10</sup> Os requisitos não funcionais denotam as principais características relacionadas às qualidades não funcionais e que atende às necessidades do cliente com eficiência.

Ao final da execução do algoritmo de mineração, os resultados devem ser disponibilizados ao usuário, para que o mesmo avalie os padrões gerados e verifique se os mesmos serão úteis em suas práticas de ensino.

### 5.2.1.2. Requisitos Não Funcionais

#### **RNF1 – O sistema deverá ter uma interface web.**

O sistema deverá ser construído para utilização na web. Assim, a sua interface deve ser apresentada num *browser*<sup>11</sup> composto de *links*, formulários e botões de ação.

#### **RNF2 – O sistema deverá ser de fácil manutenção.**

O sistema deverá utilizar uma arquitetura flexível que necessite de poucas atualizações quando mudanças nas regras de negócio existirem. Para isso, deve-se utilizar o padrão MVC (*Model-View-Controller*) (Krasner & Pope, 1988). No padrão MVC existem o objeto *Model*, que é responsável pela lógica de negócio da aplicação; o objeto *View*, responsável pela interface com o usuário; e o objeto *Controller* que define a maneira como a interface do usuário reage às entradas do mesmo.

#### **RNF3 – A interface deve ser de fácil utilização.**

Na preparação da consulta, não deverá ser exigido conhecimento de linguagens de acesso a banco de dados. Assim, o formulário de geração da consulta deverá ser de fácil manuseio, de forma a simplificar a interação do usuário com o sistema.

### 5.2.1.3. Identificação dos Atores do Sistema

Paralelamente ao levantamento dos requisitos, são identificados os atores<sup>12</sup> que utilizarão o sistema. Como os atores são identificados pelo papel que desempenham, foram identificados os seguintes papéis:

---

<sup>11</sup> Software utilizado para exibição e navegação de páginas web.

<sup>12</sup> Ator é qualquer entidade externa ao sistema que interage com o mesmo. Um ator estimula o sistema com eventos de entrada ou recebe algum resultado de processamento do mesmo (Larman, 2000).

- **Professor:** corresponde a qualquer professor da instituição de ensino, e é responsável pelo acompanhamento dos alunos das turmas nas quais o mesmo leciona;
- **Diretoria:** funcionário da instituição de ensino responsável pelo acompanhamento do desempenho dos professores em relação às disciplinas/turmas lecionadas, dos desempenhos obtidos nos cursos pelos alunos e do desempenho da instituição de ensino como um todo.

#### 5.2.1.4. Diagrama de Use-case

Um *use-case*, ou no português, caso de uso, é um documento narrativo que descreve a seqüência típica de eventos de um ator usando um sistema para completar um processo (Jacobson, *et al.*, 1992). Com use-cases, buscamos entender melhor como deverá ocorrer a interação do usuário com o sistema.

Apresentaremos os casos de uso do nosso sistema de forma estendida, que, segundo Larman (Larman, 2000), apresentam maiores detalhes do processo em questão.

A Figura 5.7 ilustra o diagrama de use-case do Midas-Poeta seguido de suas descrições.

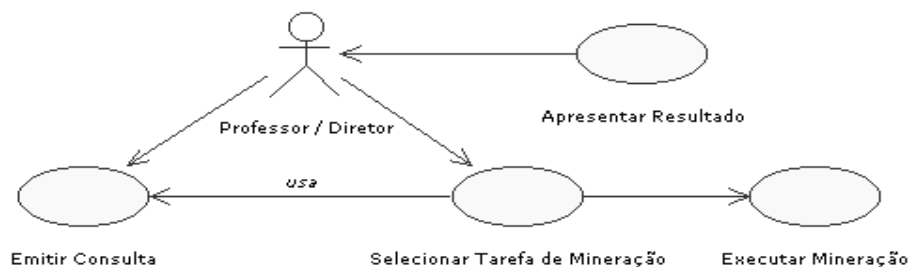


Figura 5.7. Diagrama de Use-Case do Sistema Midas-Poeta

**Use case:** *Emitir Consulta*

**Ator(es):** Professor/Diretor

**Descrição:** Um usuário aciona o sistema, prepara e emite uma consulta

#### Seqüência Típica de Eventos

Ação do ator	Resposta do sistema
--------------	---------------------

1. Acionar o sistema Midas-Poeta	2. Solicitar informações de <i>logon</i>
3. Fornecer nome/senha	4. Autenticar o usuário e disponibilizar um menu de opções de consulta
5. Escolher um consulta disponível no menu (a depender se seu papel – ver seção anterior)	6. Disponibilizar o MAD em modo gráfico
7. Selecionar no MAD as informações de interesse e submeter uma consulta ao sistema	8. Transformar a consulta do usuário em instruções SQL e executar a consulta
	9. Gerar um conjunto-resposta para a consulta

**Use case:** *Selecionar Tarefa de Mineração*

**Ator(es):** Professor/Diretor

**Descrição:** O usuário escolhe a tarefa de mineração a ser executada

#### Seqüência Típica de Eventos

<b>Ação do ator</b>	<b>Resposta do sistema</b>
1. Preparar e emitir uma consulta	2. Disponibilizar um menu de opções com as tarefas de mineração que podem ser realizadas
3. Escolher uma das tarefas de mineração disponível: associação, classificação ou agrupamento	4. Disponibilizar num formulário as medidas de interesse que podem ser aplicadas à tarefa de mineração escolhida pelo usuário
5. Informar os valores das medidas de interesse disponíveis e confirmar a mineração	6. Preparar o conjunto-resposta transformando-o numa entrada apropriada ao algoritmo de mineração escolhido pelo usuário

#### Seqüência Alternativa de Eventos

<b>Ação do ator</b>	<b>Resposta do sistema</b>
---------------------	----------------------------

1. Preparar e emitir uma consulta	2. Disponibilizar um menu de opções com as tarefas de mineração que podem ser realizadas
3. Escolher uma das tarefas de mineração disponível: associação, classificação ou agrupamento	4. Preparar o conjunto-resposta transformando-o numa entrada apropriada ao algoritmo de mineração escolhido pelo usuário

**Use case:** *Executar Mineração*

**Ator(es):** Use case “Selecionar Tarefa de Mineração”

**Descrição:** O sistema executa um algoritmo de mineração de dados

#### Seqüência Típica de eventos

Resposta do sistema
1. Executar o algoritmo de indução de conhecimento especificado segundo a seleção da tarefa de mineração e utilizar o conjunto de dados preparado para tal mineração

**Use case:** *Apresentar Resultado*

**Ator(es):** Sistema

**Descrição:** O resultado final da mineração é apresentado ao usuário

#### Seqüência Típica de Eventos

Resposta do sistema
1. Apresentar ao usuário o resultado final da mineração: regras de associação, se a tarefa escolhida foi associação; um classificador, se a tarefa de mineração escolhida foi a de classificação; ou agrupamentos, se a mineração foi baseada na tarefa de <i>clustering</i>

### 5.2.2. Projeto do Sistema Midas-Poeta

Nesta seção, detalhamos os aspectos de projeto do Midas-Poeta. Primeiro, aplicamos a estratégia descrita no capítulo 4 para obtenção do MAD. Seguindo, apresentamos a arquitetura do sistema e o diagrama de classes do sistema.

### 5.2.2.1. Fase de Planejamento do Midas-Poeta

Antes de projetarmos a arquitetura do sistema, foi necessário tomarmos algumas decisões de projeto para que o mesmo pudesse ser concebido. Tais decisões foram baseadas na fase de planejamento da estratégia de acompanhamento descrita no capítulo anterior.

#### **Definindo as Permissões dos Usuários do Sistema**

Como mostrado na seção 5.2.1.3, os usuários do sistema têm o perfil de *professor* e *diretor*. Em se tratando do perfil *professor*, o mesmo tem permissão para acessar os dados de acompanhamento de todas as suas disciplinas. Para o perfil *diretor*, o mesmo tem acesso aos dados de acompanhamento de toda a instituição.

#### **Definindo e Modelando os Dados de Interesse**

Analisando o cenário, percebemos que existe um sistema de EAD com base de dados, porém a mesma não dá suporte ao acompanhamento do aprendizado com *Data Mining*. Sendo assim, devemos definir os padrões de consulta para os perfis professor e diretor, e na seqüência, definir o MAD.

##### **Definindo os Padrões de Consulta**

O perfil professor adota os seguintes padrões de consulta:

- **Consultar uma disciplina:** o professor tem acesso aos dados de acompanhamento das turmas referentes a alguma disciplina que o mesmo leciona;
- **Consultar todas as suas disciplinas:** o professor tem acesso aos dados de acompanhamento de todas as turmas referentes a todas as disciplinas que o mesmo leciona.

Já o perfil diretor tem disponíveis as seguintes funções:

- **Consultar uma disciplina:** o diretor tem acesso aos dados de acompanhamento de todas as turmas referentes a uma disciplina em particular, não importando quem são os professores responsáveis;



- **Consultar um curso:** o diretor tem acesso aos dados de acompanhamento de todas as turmas de um determinado curso;
- **Consultar a instituição:** o diretor tem acesso aos dados de acompanhamento de todas as turmas da instituição, não importando de qual curso e disciplina a turma faz parte.

### **Definindo os Dados de Interesse e Modelagem do MAD**

A análise sobre a relevância das informações do banco de dados do Portfolio-Tutor (ver seção 5.1) mostrou que muitos atributos e entidades não são necessariamente interessantes pois armazenam dados sobre controle de acesso e segurança, dados textuais, dados sobre a programação de cursos e dados de controle de zonas de comportamento e perfis de grupos disponíveis para o funcionamento do tutor.

Ao definir os dados de interesse, devemos ter em mente que para a concepção do MAD, o sistema deverá dispor dados sobre o acompanhamento informal (requisito R4), dados sobre o acompanhamento alternativo (requisito R5) e dados demográficos, comportamentais e dados sobre avaliação (requisito R6).

Para contemplarmos os aspectos informais do acompanhamento, necessitamos de dados sobre as interações do aluno com outros alunos e sobre as interações do aluno com os materiais didáticos disponíveis. A base de dados Portfolio-Tutor prevê o armazenamento das interações do aluno com seus colegas/professores através de e-mails, estando disponível na tabela *ptuEmail*; e também prevê a interação do aluno com os recursos didáticos através da tabela *ptuHistoricoNavegacao*. Portanto, o requisito R4 da estratégia está atendido.

Para contemplar os aspectos alternativos do acompanhamento, precisamos de dados sobre os itens de julgamento da avaliação autêntica. Neste sentido, o sistema Portfolio-Tutor também oferece subsídios, pois o mesmo já foi projetado idealizando a avaliação autêntica. Assim, a sua base de dados prevê os dados necessários ao acompanhamento alternativo através da entidade *ptuAvaliacaoItemAvaliacao*, que armazenam os resultados dos itens de avaliação para cada artefato produzido pelo aluno, e das entidades *ptuArtefato* e *ptuItemAvaliacao*, que armazenam os artefatos e itens de avaliação padronizados pela instituição, respectivamente.

Para finalizar a definição dos dados de interesse, falta a definição dos dados demográficos, dados sobre avaliação e dados comportamentais do aluno. Com relação aos dados demográficos e comportamentais, o Portfolio-Tutor não apresenta nenhum recurso pelo qual se possa capturar estes dados. Apenas se utiliza um cadastro de aluno contendo dados sobre o ano de conclusão do curso que o mesmo frequenta e o seu coeficiente de rendimento escolar (disponível na tabela *ptuAluno*). Já para os dados sobre avaliação, o Portfolio-Tutor apresenta uma tabela que armazena as atividades realizadas pelo aluno, chamada *ptuAtividade*, e uma tabela que armazena o desempenho em cada atividade, chamada *ptuAtividadeAluno*. Porém, o mesmo não oferece a disponibilização de dados por tipos de avaliações. Sendo assim, o requisito R6 não é satisfeito no Portfolio-Tutor, e de acordo com a estratégia, deverá ser incrementado.

No cadastro do aluno devem ser acrescentados dados sobre a região de procedência, idade, sexo (dados demográficos) e a situação do aluno no curso (dado comportamental). Para os dados sobre as avaliações, necessitamos criar uma tabela de tipos de avaliação, e acrescentar a sua chave como chave estrangeira na tabela de atividades, para que possamos nos referir a uma atividade como sendo uma prova, ou exercício, ou seminário, etc. A Figura 5.8 ilustra a alteração no esquema lógico do banco de dados do Portfolio-Tutor.

```

- Alter table ptuAluno add sexo char(1) not null check (sexo in ('F','M'))
- Alter table ptuAluno add uf char(2) not null check (uf in ('BA','CE','PB','PE','SP'))
- Alter table ptuAluno add idade smallint not null check (idade between 10 and 80)
- Alter table ptuAluno add situacao char(5) not null check (situacao in ('CURSA','DEISIS','TRANC'))
- Create table ptuTipoAtividade ( idTipoAtiv int not null primary key, dsTipoAtividade varchar(20))
- Alter table ptuAtividade add idTipoAtiv int not null references ptuTipoAtividade (idTipoAtiv)

```

Figura 5.8. Alteração do Esquema Lógico do Banco de Dados do Portfolio-Tutor

Como um MAD reflete as informações de interesse relacionado a um aluno, o modelamos, para o sistema Midas-Poeta, da seguinte forma:

<p>▪ <b>DADOS DO ACOMPANHAMENTO INFORMAL</b></p>
<p><b>Interação por E-mail:</b> número de mensagens de e-mail enviado por um aluno numa disciplina de um professor</p>

<b><u>Interação por Recurso Didático:</u></b> número médio de acesso do aluno aos recursos didáticos disponíveis numa disciplina de um professor
▪ <b><i>DADOS DO ACOMPANHAMENTO ALTERNATIVO</i></b>
<b><u>Aproveitamento por Item de Avaliação:</u></b> desempenho médio do aluno num item de avaliação utilizado pelo professor numa disciplina
<b><u>Aproveitamento por Artefato:</u></b> desempenho médio do aluno por tipo de artefato utilizado pelo professor numa disciplina
▪ <b><i>DADOS SOBRE AVALIAÇÃO</i></b>
<b><u>Aproveitamento por Tipo de Atividade:</u></b> desempenho médio do aluno por tipo de avaliação feita pelo professor numa disciplina
▪ <b><i>DADOS DEMOGRÁFICOS</i></b>
<b><u>Região de Procedência:</u></b> representa o estado (UF) que o aluno nasceu
<b><u>Sexo:</u></b> representa o sexo do aluno
<b><u>Idade:</u></b> representa a idade do aluno
▪ <b><i>DADOS COMPORTAMENTAIS</i></b>
<b><u>Situação:</u></b> representa a situação do aluno no curso (desistente, trancado, cursando)

Figura 5.9. Modelo Analítico de Dados do Midas-Poeta

### Implementação do MAD

A fim de implementarmos o MAD acima, foi utilizada a abordagem virtual como mecanismo de visões. Sendo assim, foi definida para cada informação de interesse uma visão correspondente. A Figura 5.10 abaixo ilustra o esquema de visões.

<b><i>Interação por E-mail</i></b> (idAluno, idDisciplina, idProfessor, <b><u>numMensagens</u></b> )
<b><i>Interação por Material Didático</i></b> (idAluno, idDisciplina, idProfessor, idMaterial, <b><u>numAcessos</u></b> )
<b><i>Aproveitamento por Item de Avaliação</i></b> (idAluno, idDisciplina, idProfessor, idItem, <b><u>desempenho</u></b> )
<b><i>Aproveitamento por Artefato</i></b> (idAluno, idDisciplina, idProfessor, idArtefato, <b><u>desempenho</u></b> )
<b><i>Aproveitamento por Atividade</i></b> (idAluno, idDisciplina, idProfessor, idAtividade, <b><u>desempenho</u></b> )
<b><i>Região de Procedência</i></b> (idAluno, <b><u>uf</u></b> )
<b><i>Sexo</i></b> (idAluno, <b><u>sexo</u></b> )
<b><i>Idade</i></b> (idAluno, <b><u>idade</u></b> )
<b><i>Situação</i></b> (idAluno, <b><u>situação</u></b> )

Figura 5.10. Conjunto de Visões do Modelo Analítico de Dados

## Disponibilidade de Tarefas de Mineração

Para finalizar a parte de planejamento, deveremos ter em mente quais tarefas de mineração serão disponíveis aos usuários do sistema. Para o Midas-Poeta, dispomos as funções de associação, classificação e agrupamento. A razão desta escolha foi influenciada pelas idéias de Linderman (Linderman, 1986), onde o mesmo atribuiu, como resultado da tomada de decisão na educação presencial, operações preditivas e classificatórias (rever seção 3.2.3). Assim, consideramos a sua teoria como base também para o nosso sistema.

Com a tarefa de associação, pode-se encontrar padrões úteis nas correlações de dados de diversas categorias (dados informais *versus* dados alternativos, dados demográficos *versus* dados sobre avaliação, etc). Com as tarefas de classificação e agrupamento, os problemas citados por Linderman na operação classificatória podem ser desconsiderados, já que não há nenhuma intervenção humana para realizar tal tarefa.

### 5.2.2.2. Arquitetura do Sistema Midas-Poeta

A arquitetura define as partes que compõe o sistema e o relacionamento existentes entre elas. Um modelo usualmente utilizado é a arquitetura em três camadas que separa a lógica da aplicação em uma camada intermediária, entre as camadas de *interface* com o usuário e o armazenamento persistente dos dados (Larman, 2000).

Considerando esta abordagem e utilizando a representação de pacotes<sup>13</sup>, a Figura 5.11 apresenta a arquitetura do Midas-Poeta. Neste caso, a lógica da aplicação foi decomposta em duas camadas: a Camada de Domínio e Camada de Serviços. A Camada de Domínio é constituída dos pacotes Menu, DataModel e DataMining. A Camada de Serviços é composta pelo pacote de Acesso a Banco de Dados.

---

<sup>13</sup> Pacotes são mecanismos utilizados pela UML (*Unified Modeling Language*) para o agrupamento de elementos semanticamente interligados.

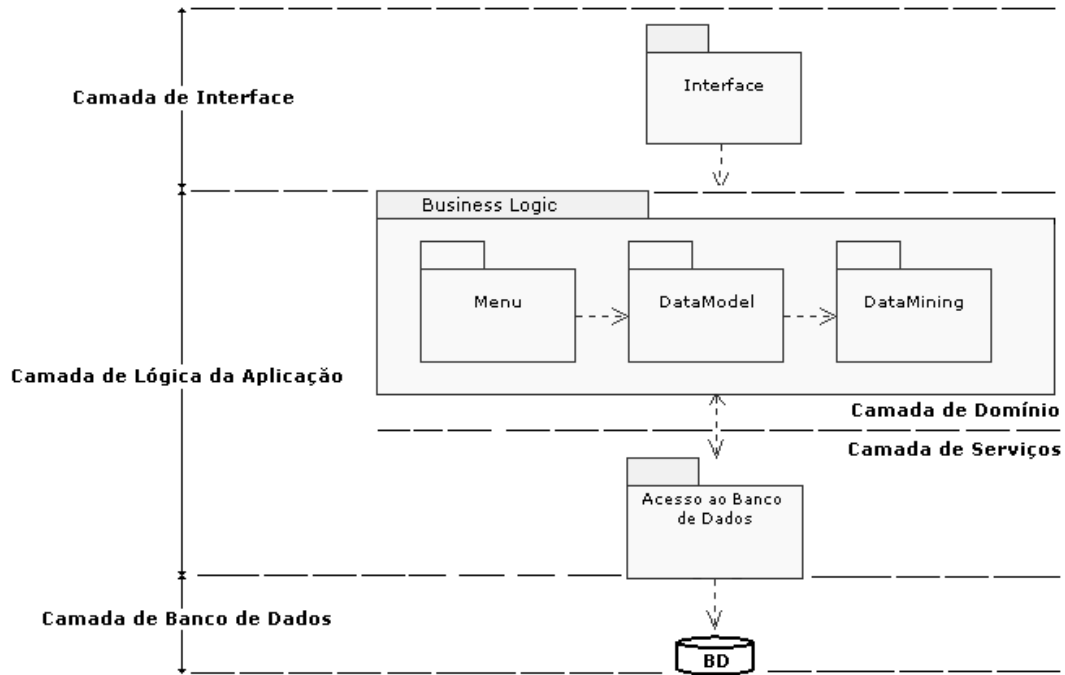


Figura 5.11. Arquitetura do Midas-Poeta

A Camada de Banco de Dados corresponde ao banco de dados relacional *Microsoft SQL Server 7.0*, onde o conjunto de tabelas e visões manipuladas na Camada de Domínio é armazenado.

A Camada de Serviços é composto pelo pacote de Acesso a Banco de Dados, que é responsável pela comunicação da aplicação com o banco de dados.

Na Camada de Domínio, temos o pacote Menu, que corresponde à classe que disponibiliza o menu de acordo com o perfil do usuário; o pacote DataModel, que corresponde às classes que representam os conceitos do domínio e modelam a interação do usuário com o Modelo Analítico de Dados; e por fim, o pacote DataMining, que contém o conjunto de classes que ativam os algoritmos de mineração de dados.

A Figura 5.12 ilustra o diagrama de classes<sup>14</sup>, na notação UML (*Unified Modeling Language*) da arquitetura do Midas-Poeta.

<sup>14</sup> O diagrama de classes é utilizado para fazer a modelagem da visão estática do projeto de um sistema, ele apresenta o conjunto de classes, interfaces e seus relacionamentos (Booch, *et al.*, 2000).

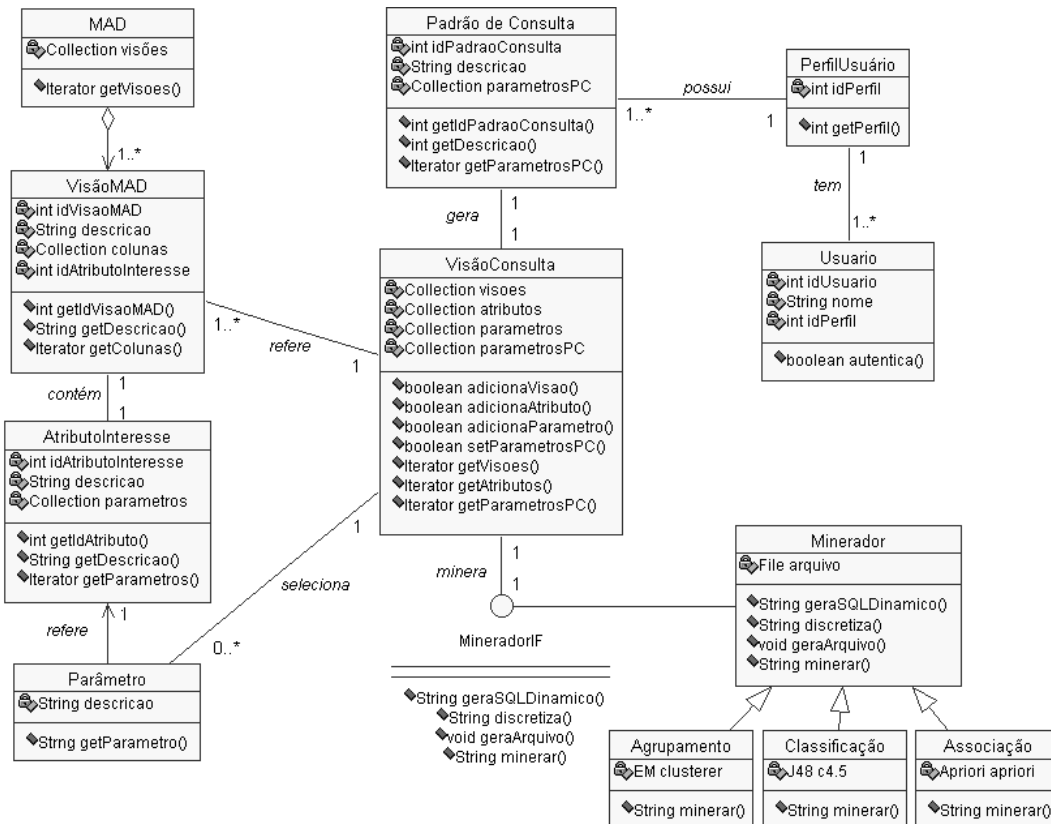


Figura 5.12. Diagrama de Classes da Arquitetura do Midas-Poeta

Descrevemos as principais classes do diagrama:

- **Padrão de Consulta:** representa os tipos de consultas de um perfil de usuário. Dentre os seus atributos, podemos destacar o atributo “parametrosPC”, que representa o conjunto de parâmetros de um padrão de consulta. Por exemplo, se um professor realizar uma consulta “Por Disciplina”, os valores do código do professor e do código da disciplina devem ser guardados na coleção;
- **VisãoMAD:** representa uma visão disponível no MAD. Dentre os seus atributos, podemos destacar o atributo “colunas”, que representa as colunas da visão implementada no banco de dados, e o atributo “idAtributoInteresse”, que é a coluna que representa o atributo de interesse (por exemplo, nº de mensagens, idade, etc);
- **Parâmetro:** Representa um valor associado ao contexto de uma visão do MAD e que é utilizado pelo usuário para selecionar parte dos dados daquela visão. Por exemplo, para a visão “Aproveitamento por Item de Avaliação”, o usuário

poderia selecionar apenas o aproveitamento para os itens de avaliação “criatividade”, “capacidade de liderança” e “capacidade crítica”;

- **VisãoConsulta:** representa uma coleta de dados realizada pelo usuário. Possui os atributos “visões”, que são as visões selecionadas pelo usuário, “atributos”, que são os atributos de interesse de cada visão, “parâmetros”, que são os parâmetros selecionados para restringir a consulta, e “parametrosPC”, que são os parâmetros do padrão de consulta do usuário;
- **Minerador:** classe abstrata que implementa a interface “MineradorIF”. É abstrata pois o método “minerar” é realizado nas subclasses “Agrupamento”, “Associação” e “Classificação”. Para isso, nestas subclasses existe um atributo que corresponde a uma classe que implementa um algoritmo de mineração de dados. Outro método que a classe abstrata “Minerador” implementa e que merece um comentário é o método “geraSQLDinamico”. Tal método é realizado utilizando as coleções de “visões”, “atributos”, “parâmetros” e “parâmetrosPC” da classe “VisãoConsulta” para gerar um comando SQL dinâmico de junções entre as *views* do banco de dados que representam as visões selecionadas no MAD.

O comando SQL é montado da seguinte forma:

```

Select <Visões.Atributos>
From <Visões>
Where
    <Parâmetros> and
    <ParâmetrosPC>
    
```

Estes comandos SQL geram como resultado, uma tabela contendo os atributos da seleção. A Tabela 5.1 ilustra um exemplo de um resultado parcial genérico.

Atributo1	Atributo2	Atributo3	AtributoN
Valor11	Valor21	Valor31	ValorN1
Valor12	Valor22	Valor32	ValorN2
.....	.....	.....	.....

**Tabela 5.1.** Exemplo de um Resultado Parcial do SQL Dinâmico

O método “discretiza” realiza a discretização dos valores do resultado parcial, transformando-o num conjunto de dados preparado para a mineração. Para isso, é utilizada uma tabela de discretização, que associa uma visão do MAD a um conjunto de valores discretizados. Um subconjunto da tabela de mapeamento é exemplificado na Tabela 5.2. A tabela de discretização completa pode ser consultada no Anexo 1.

<b>VISÃO</b>	<b>FAIXA DE VALORES</b>	<b>VALOR DISCRETIZADO</b>
Idade	Menor que 21	Jovem
Idade	21 .. 30	Adulto
Idade	31 .. 45	Maduro
Idade	Maior que 45	Velho
Interação p/ Material Didático	0	Nenhum
Interação p/ Material Didático	1 .. 3	Pouco Freqüente
Interação p/ Material Didático	4 .. 5	Regular
Interação p/ Material Didático	6 .. 7	Freqüente
Interação p/ Material Didático	Maior que 7	Muito Freqüente

**Tabela 5.2.** Tabela de Mapeamento para Discretização

Por fim, o método “geraArquivo” recebe o conjunto de dados preparado e discretizado e cria um arquivo de dados que será aplicado ao algoritmo de mineração.

### 5.2.3. Implementação do Midas-Poeta

Esta seção abrange os principais aspectos relacionados à implementação do sistema Midas-Poeta. Apresentaremos os aspectos relevantes às tecnologias empregadas, a implementação da fase de análise de dados, a disposição física dos módulos do sistema e as interfaces disponibilizadas.

#### 5.2.3.1. Tecnologias Empregadas

O sistema Midas-Poeta foi desenvolvido como aplicação *Web*, onde o acesso às informações se dá a partir de um *browser*, sendo portanto desnecessária a instalação de qualquer software adicional na estação dos usuários finais.



A lógica de negócio da aplicação foi implementada utilizando a linguagem de programação Java; e para a disponibilização dinâmica do conteúdo e a criação das interfaces, foram empregadas as tecnologias JSP (*Java Server Pages*) e HTML.

A linguagem JSP é fundamentada na linguagem Java e sua grande vantagem é que a mesma simplifica o processo de desenvolvimento de aplicações *Web*, já que possibilita a separação entre a lógica da aplicação (código Java) e a apresentação das informações. *Java Server Pages* são arquivos de texto (com extensão “.jsp”), que contém HTML tradicional junto com código Java embutido, através da utilização de marcações (*tags*) JSP, de forma que o código da página é executado todas as vezes que uma solicitação for recebida pelo servidor *Web*, e o conteúdo dinâmico é gerado e unido ao conteúdo estático antes de ser enviado ao usuário (Fields & Kolb, 2000).

Como Java é a linguagem básica utilizada na tecnologia JSP, esta última passa então a usufruir também dos benefícios provenientes da primeira, dentre as quais pode-se citar (Fields & Kolb, 2000):

- Código mais robusto, pelo emprego de tratamento de exceções e gerenciamento automático de memória;
- Portabilidade entre as diversas plataformas de hardware e software;
- Vários recursos disponíveis, inclusive os que permitem o acesso à banco de dados, processamento distribuído, criptografia, dentre outras.

Para implementarmos o conjunto de visões do MAD, foi utilizado o sistema gerenciador de banco de dados relacional *Microsoft SQL Server 7.0*. A aplicação acessa o banco de dados através da tecnologia JDBC (*Java Database Connectivity*). Esta tecnologia corresponde a uma API, que permite que programas escritos na linguagem Java possam acessar os dados armazenados em uma grande variedade de sistemas gerenciadores de banco de dados relacionais, desde que seja fornecido o *driver* específico. Através da API, o programa Java pode conectar-se ao banco de dados e executar comandos SQL.

### 5.2.3.2. Implementação da Fase de Análise de Dados

Para atender a etapa de coleta de dados, utilizamos a tática de criação de comandos SQL dinâmicos, onde junções são realizadas entre as visões, atributos e parâmetros selecionados pelo usuário gerando um *resultset*, que é preparado (limpeza, discretização e transformação) antes da aplicação de um algoritmo de *Data Mining*. A limpeza é garantida pelo próprio comando SQL que realiza a junção. A discretização é baseada em valores pré-estabelecidos segundo a tabela de discretização, e a transformação é realizada utilizando o método “geraArquivo” da classe de “Minerador” implementada no pacote DataMining (ver seção 5.2.2.2).

Na implementação das tarefas de mineração de dados, foi utilizado o *framework* WEKA (Witten & Frank, 2000). Tal ambiente provê uma série de algoritmos de mineração de dados, todos escritos em Java. No Midas-Poeta, utilizamos os algoritmos C4.5 (Quinlan, 1993), Apriori (Agrawal & Srikant, 1994) e EM (Lauritzen, 1995) para realizarem as tarefas de classificação, associação e agrupamento, respectivamente.

Para visualização dos resultados da mineração, preferimos exibir os resultados da mineração produzidos pelo próprio WEKA. Sugere-se que, futuramente, essas regras sejam exibidas ao usuário em um formato mais compreensível, possivelmente gráfico.

### 5.2.3.3. Disposição Física dos Módulos do Sistema

A tecnologia empregada para o desenvolvimento do sistema Midas-Poeta possibilita a disposição física dos seus módulos em três camadas (Figura 5.13), onde a primeira camada, a interface da aplicação com o usuário, é feita por um navegador executado na estação de trabalho; a camada central, que corresponde à aplicação, é executada por um servidor *Web* (estendido com um container JSP) e, enfim, a terceira camada, o armazenamento, é realizado pelo servidor de banco de dados.

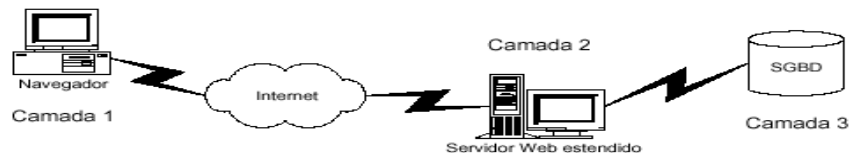


Figura 5.13. Disposição Física em Três Camadas

#### 5.2.3.4. As Interfaces do Midas-Poeta

Como o Midas-Poeta foi desenvolvido baseado em tecnologia *Web*, a interface do sistema é basicamente o *browser*, com as informações geradas pela aplicação ocupando a área existente para navegação.

Ao acionar o sistema, o usuário necessita de uma autenticação, informando **nome e senha de acesso**, conforme apresentado na Figura 5.14.

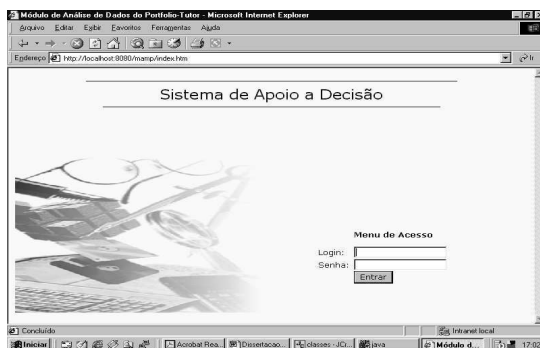


Figura 5.14. Interface de Logon

O sistema de posse dessas informações, irá validar o usuário, autenticar o usuário no banco de dados, identificar o seu perfil e direcionar para a interface respectiva ao seu perfil.

De acordo com o perfil do usuário, será apresentado um menu de opções de consultas (padrão de consulta). A seguir, a Figura 5.15 ilustra o menu para o perfil diretor e professor, respectivamente.

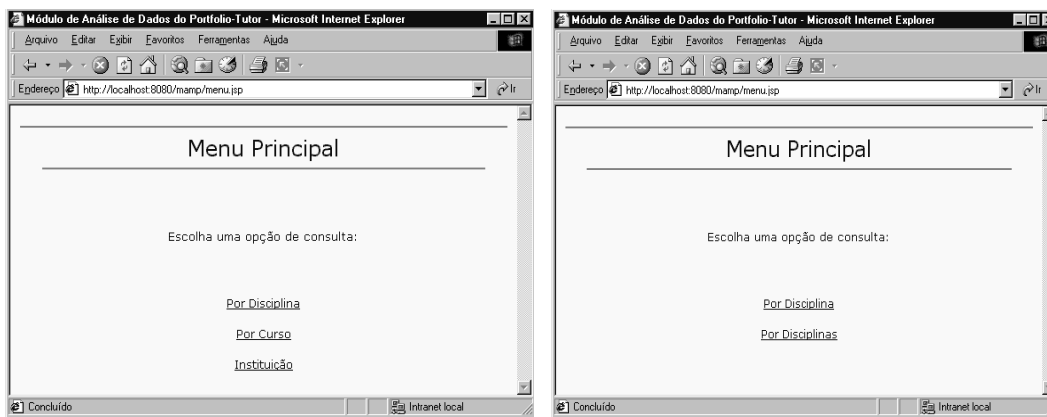


Figura 5.15. Interfaces dos Menus de Consulta Disponível para o Professor e para o Diretor

Para os perfis diretor e professor, ao ser escolhido a opção “Por Disciplina”, é apresentada a tela ilustrada na Figura 5.16, onde o usuário seleciona a disciplina na qual deseja descobrir padrões de acompanhamento.

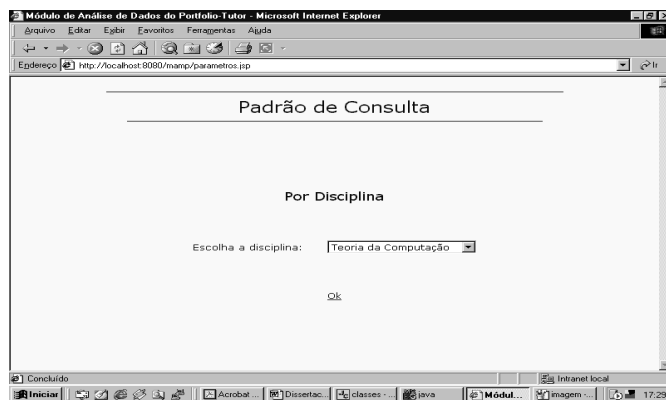


Figura 5.16. Tela do Padrão de Consulta “Por Disciplina”

Para o perfil diretor, ao ser escolhida a opção “Por Curso”, é apresentada a interface conforme a Figura 5.17.

Para os padrões de consulta “Instituição” e “Por Disciplinas” dos perfis diretor e professor, respectivamente, não existem interfaces, já que estas consultas não apresentam nenhum parâmetro.

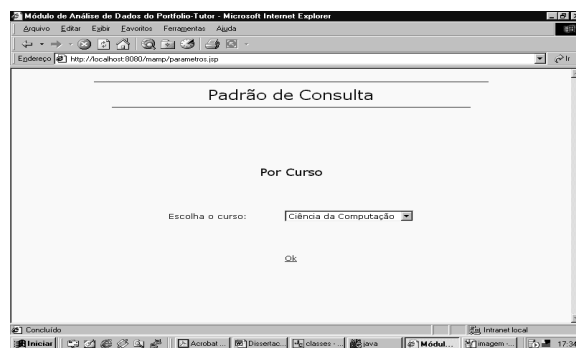


Figura 5.17. Interface do Padrão de Consulta “Por Curso”

Determinado o padrão de consulta, a próxima interface é a de coleta de dados. Conforme a Figura 5.18, são apresentadas as visões e os atributos de interesse

disponíveis no Modelo Analítico de Dados do Midas-Poeta para a realização dessa coleta.

Etapa de Coleta de Dados

VISÕES

**Interação por E-Mail**  
 Nº de Mensagens

**Interação por Tipo de Material Didático**  
 Nº de Acessos ..... Todos

**Aproveitamento por Tipo de Atividade**  
 Rendimento ..... Seminário

**Aproveitamento por Item de Avaliação**  
(1)  Rendimento ..... (2) Clareza de Conteúdo

**Aproveitamento por Artefato**  
 Rendimento ..... Todos

**Aluno**  
 Sexo ..... Todos  
 Idade .....  
 Estado ..... Todos  
 Situação ..... Todos

Avançar

Figura 5.18. Interface de Coleta de Dados

Ao selecionar um atributo de interesse, o usuário marca o *checkbox* (1) correspondente. Para restringir a pesquisa, o usuário tem a possibilidade de escolher quais valores daquele atributo ele deseja incluir em sua consulta usando o *listbox* (2) ao lado do nome do atributo. Por exemplo, poderia ser escolher o atributo “Rendimento” da visão “Aproveitamento por Item de Avaliação”, porém, deve-se consultar apenas o rendimento do item “Clareza de Conteúdo”. No *listbox*, o usuário tem a possibilidade de escolher um valor ou mais de um valor. Para simplificar a escolha de todos os valores disponíveis, foi incluído em cada *listbox* o valor “TODOS”, no qual seleciona todos os valores automaticamente.

Para a etapa de mineração de dados, foi desenvolvida a interface apresentada na Figura 5.19. O usuário tem a possibilidade de escolher as tarefas de mineração: **agrupamento**, que não apresenta nenhuma medida de interesse; **associação**, que possui as medidas suporte e confiança; e **classificação**, onde o usuário deve informar o atributo de classificação e a confiança, que expressa o grau de confiança da acurácia do conjunto de execução (Toebe, 2002).

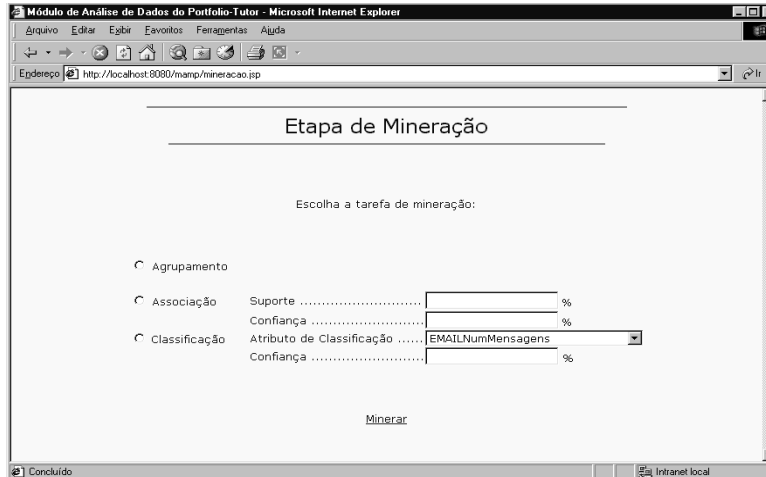


Figura 5.19. Interface da Etapa de Mineração

Por fim, a última tela do sistema Midas-Poeta é a de avaliação e visualização dos resultados. Conforme a Figura 5.20, tal interface é composta por uma área de texto no qual é exibido o padrão de conhecimento produzido por algum algoritmo de mineração.



Figura 5.20. Interface da Etapa de Avaliação e Visualização dos Resultados

### 5.3. Considerações Finais

Este capítulo apresentou os aspectos relacionados ao desenvolvimento do Midas-Poeta. Tal sistema foi concebido como um módulo de análise de dados para o ambiente Portfolio-Tutor e emprega a estratégia de acompanhamento do aprendizado descrita no capítulo anterior.

Com a utilização do Midas-Poeta, professores e diretores têm a possibilidade de acompanhar os estudantes de maneira mais elaborada, onde suas decisões são apoiadas por um conjunto de informações de acompanhamento e por um módulo de geração semi-automática de padrões de conhecimento sobre a aprendizagem.

### *UM CENÁRIO DE EXECUÇÃO DO MIDAS-POETA*

Este capítulo apresenta um estudo de caso ilustrando o funcionamento do Midas-Poeta. Inicialmente são feitas algumas considerações a cerca do cenário de execução. Na seqüência, apresentaremos as interações de um professor e do diretor com o sistema. Finalizamos com algumas considerações.

#### 6.1. Considerações Iniciais

Para o estudo de caso utilizamos dados fictícios de um curso fictício de pós-graduação em macroeconomia à distância.

O curso contém duas disciplinas: “Macroeconomia Geral” e “Macroeconomia Regional”. Cada uma das disciplinas é lecionada por um professor diferente, que está alocado nela desde a criação do curso.

Como a instituição adotou o sistema Portfolio-Tutor para gerenciar o curso à distância, ela segue o paradigma de avaliação autêntica com o uso de um portfolio.

A escola padronizou os seguintes artefatos: *Projeto de Pesquisa*, *Sumário*, *Crítica de Artigo* e *Experiência em Laboratório*. Os artefatos serão avaliados segundo os itens de avaliação *Criatividade*, *Vocabulário Diversificado*, *Capacidade Crítica* e *Clareza de Conteúdo*.

Os tipos de recursos didáticos utilizados no curso podem ser: *Conceito*, *Exercício*, *Exemplo* e *Roteiro*. E os possíveis tipos de atividade são: *Seminário* (via Videoconferência), *Prova*, *Exercício*, *Mini-Teste* e *Questionário*.

O povoamento da base de dados foi realizado em duas etapas. Na primeira, foram inseridos manualmente os dados relativos ao curso, professores, disciplinas, artefatos e itens de avaliação padronizados pela instituição de ensino, tipos de atividade e os tipos de recursos didáticos. Foi realizada também a alocação de um professor na



disciplina “Macroeconomia Geral” e a alocação de outro professor na disciplina “Macroeconomia Regional” em dez períodos (5 anos).

A segunda etapa do povoamento se refere ao funcionamento do curso, ou seja, criação de turmas, registro das interações dos alunos aos materiais didáticos, registro das interações dos alunos através de mensagens de e-mail, atribuição de atividades e notas, resultado dos julgamentos dos itens de avaliação feitos aos artefatos, atualização da situação dos alunos no curso e o próprio povoamento dos dados cadastrais dos alunos (idade, sexo, região de procedência e outros).

As atribuições de atividades e artefatos, e os tipos de materiais didáticos disponibilizados pelo professor em cada período foram feitos manualmente para cada disciplina.

Para a criação das turmas, preferimos cadastrar 50 novos alunos em cada período, onde os dados cadastrais dos alunos foram aleatoriamente inseridos na base de dados.

Quando as turmas são criadas, os dados referentes as interações por materiais didáticos, interações por e-mail, atribuições de notas e resultados dos julgamentos de itens de avaliação feitos aos artefatos são inseridos aleatoriamente no banco de dados. Para isso, foi criado um programa Java que lê as turmas e armazenava para cada aluno os valores de tais dados. Adicionalmente, o programa atualiza a situação do aluno no curso (cursando/trancado/desistente) também de forma randômica.

Ao final deste processo de carga de base de dados, em cada disciplina, foram matriculados 500 alunos. Portanto, 1000 alunos foram matriculados no curso de pós-graduação em macroeconomia da instituição de ensino.

## 6.2. Interagindo com o Midas-Poeta

A utilização do Midas-Poeta pode ser realizada a qualquer momento do processo de ensino/aprendizagem. Porém, é de valia notar algumas situações interessantes:

- O professor pode utilizar o sistema no final de um semestre letivo ou no início do semestre. Assim, ele pode averiguar o aprendizado passado e presente e realizar previsões sobre o aprendizado futuro;

- O professor pode utilizar o sistema quando existir um baixo desempenho acadêmico dos alunos. Desta forma, ele pode detectar problemas no ensino e corrigi-los em tempo hábil, sem que a(s) sua(s) turma(s) seja(m) prejudicada(s);
- O professor pode utilizar o sistema ao final de um período letivo realizando uma consulta já explorada em períodos anteriores. Assim, ele pode verificar se houve alguma mudança entre o padrão de conhecimento passado e o padrão de conhecimento gerado no presente. Esta situação é importante, já que à medida que novos dados vão sendo inseridos na base de dados, novos padrões de conhecimento podem ser descobertos, tornando o antigo padrão ultrapassado ou não mais útil;
- O diretor pode usar o sistema no final de um semestre letivo para averiguar problemas de ensino em uma ou mais disciplinas. Com isso, o diretor pode, em reuniões programadas, expor tais problemas aos professores de forma que eles possam tentar solucioná-los;
- O diretor pode utilizar o sistema quando detectar um alto índice de trancamentos ou desistência. Assim, ele pode identificar o motivo de tais abandonos.

A seguir, apresentaremos uma interação de um professor e uma interação do diretor do curso fictício. Para tal, a seleção dos atributos para a mineração foi realizada sem que nenhuma hipótese fosse sugerida, já que a base de dados do sistema Portfolio-Tutor foi povoada com dados aleatórios.

### 6.2.1. A Interação de um Professor com o Midas-Poeta

O professor da disciplina “Macroeconomia Geral” desejou correlacionar o aproveitamento dos itens de avaliação “Criatividade” e “Capacidade Crítica” com o número de acessos aos materiais didáticos tipo “Conceito” e a situação do aluno no curso. Para isso, foram selecionados os atributos e valores das seguintes visões: (1) **“rendimento”** dos itens de avaliação “Criatividade” e “Capacidade Crítica” da visão **Aproveitamento por Item de Avaliação**; (2) **“nº de Acessos”** ao material didático “Conceito” da visão **Interação por Material Didático**; e (3) **“situação”** da visão **Situação**. A Figura 6.1 ilustra tal seleção.

## Etapa de Coleta de Dados

---

**VISÕES**

**Interação por E-Mail**  
 Nº de Mensagens

**Interação por Tipo de Material Didático**  
 Nº de Acessos ..... Conceito

**Aproveitamento por Tipo de Atividade**  
 Rendimento ..... Todos

**Aproveitamento por Item de Avaliação**  
 Rendimento ..... Criatividade

**Aproveitamento por Artefato**  
 Rendimento ..... Capacidade Crítica

**Aluno**  
 Sexo ..... Todos  
 Idade .....  
 Estado ..... Todos  
 Situação ..... Todos

[Avançar](#)

Figura 6.1. Etapa de Coleta de Dados Realizada pelo Professor

Com esta especificação de consulta, foi gerado o seguinte comando SQL:

```

Select
    MATERIALConceito.MATERIALConceito,
    ITEMCriatividade.ITEMCriatividade,
    ITEMCapacidadeCritica.ITEMCapacidadeCritica,
    ALUNOSituacao.ALUNOSituacao
From
    (Select idAluno,
    numAcessos as MATERIALConceito
From
    InteracaoTipoRecursoDidatico
Where
    idDisciplina=1 and
    idProfessor=1 and
    idMaterialDidatico=1) MATERIALConceito,
    (Select
    idAluno,
    AproveitamentoItemAvaliacao as ITEMCriatividade
From
    AproveitamentoItemAvaliacao
Where
    idDisciplina=1 and
    idProfessor=1 and
    idItem=1) ITEMCriatividade,
    (Select
    idAluno,
    AproveitamentoItemAvaliacao as ITEMCapacidadeCritica
From
    AproveitamentoItemAvaliacao

```

```

Where
    idDisciplina=1 and
    idProfessor=1 and idItem=3) ITEMCapacidadeCritica,
(Select
    idAluno,
    Situacao as ALUNOSituacao
From
    AlunoSituacao) ALUNOSituacao
Where
    MATERIALConceito.idAluno=ITEMCriatividade.idAluno and
    ITEMCriatividade.idAluno=ITEMCapacidadeCritica.idAluno and
    ITEMCapacidadeCritica.idAluno=ALUNOSituacao.idAluno

```

O resultado desta seleção foi posteriormente discretizado e preparado para a mineração de dados. A Tabela 6.1 apresenta parte deste resultado.

NºAcessoConceito	RendimentoCriatividade	RendimentoCapCritica	Situação
POUCO	REGULAR	ÓTIMO	CURSA
MUITA	ÓTIMO	EXCELENTE	CURSA
REGULAR	REGULAR	BOM	DESISTENTE
.....	.....	.....	.....

**Tabela 6.1.** Resultado da Coleta de Dados feita pelo Professor

A tarefa de mineração escolhida foi a de associação com suporte 60% e confiança 70%. A Figura 6.2 ilustra o resultado da mineração.



**Figura 6.2.** Resultado Mineração por Associação feita pelo Professor

Sua interpretação é a seguinte:

- 60% dos alunos com acesso “Regular” ao material “Conceito” estão na situação “Desistente”;
- 60% dos alunos com rendimento “Excelente” no item “Criatividade” e com rendimento “Ruim” no item “Capacidade Crítica” têm acesso “Pouco Frequente” ao material didático “Conceito”;
- 60% dos alunos com rendimento “Excelente” no item “Capacidade Crítica” têm acesso “Excelente” ao material didático “Conceito”, apresentam rendimento “Excelente” no item “Criatividade” e estão na situação “Trancado”.

Deve ser levado em conta que, como os dados foram gerados aleatoriamente, os resultados são completamente hipotéticos. Assim, a avaliação desse resultado é desnecessária.

### 6.2.2. A Interação de um Diretor com o Midas-Poeta

O diretor da instituição resolveu classificar um aluno como “desistente” ou “trancado” do curso correlacionando a sua região de procedência (Estado) com o desempenho em provas. Para isso, foram selecionados os atributos e valores das seguintes visões: (1) “*rendimento*” na atividade “Prova” da visão **Aproveitamento por Tipo de Atividade**; (2) “*estado*” da visão **Estado**; e (3) os valores “Desistente” e “Trancado” do atributo “*situação*” da visão **Situação**. A Figura 6.3 ilustra tal seleção.

The screenshot shows a web interface titled "Etapa de Coleta de Dados" with a section for "VISÕES". The following filters are visible:

- Interação por E-Mail**:  Nº de Mensagens
- Interação por Tipo de Material Didático**:  Nº de Acessos ..... Todos
- Aproveitamento por Tipo de Atividade**:  Rendimento ..... Prova
- Aproveitamento por Item de Avaliação**:  Rendimento ..... Todos
- Aproveitamento por Artefato**:  Rendimento ..... Todos
- Aluno**:
  - Sexo ..... Todos
  - Idade
  - Estado ..... Todos
  - Situação ..... Desistente, Trancado

An "Avançar" button is located at the bottom of the form.

Figura 6.3. Etapa de Coleta de Dados Realizada pelo Diretor

Com esta especificação de consulta, foi gerado o seguinte comando SQL:

```

Select
    ATIVIDADEProva. ATIVIDADEProva,
    ALUNOEstado. ALUNOEstado,
    ALUNOSituacao.ALUNOSituacao
From
    (Select
        idAluno,
        avg(desempenho) as ATIVIDADEProva
    From
        AproveitamentoTipoAtividade
    Where
        idTipoAtividade=2
    Group by
        idAluno) ATIVIDADEProva,
    (Select
        idAluno,
        uf as ALUNOEstado
    From
        AlunoEstado) ALUNOEstado
    (Select
        idAluno,
        Situacao as ALUNOSituacao
    From
        AlunoSituacao) ALUNOSituacao
Where
    ATIVIDADEProva.idAluno=ALUNOEstado.idAluno and
    ALUNOEstado.idAluno=ALUNOSituacao
    
```

O resultado desta seleção foi posteriormente discretizado e preparado para a mineração de dados. A Tabela 6.2 apresenta parte deste resultado.

DesempenhoProva	Estado	Situação
ÓTIMO	PB	TRANCADO
FRACO	CE	DESISTENTE
ÓTIMO	PB	TRANCADO
.....	.....	.....

**Tabela 6.2.** Resultado da Coleta de Dados feita pelo Diretor

Ao escolher a tarefa de classificação, foi informado como atributo de classificação a coluna “Situação” com confiança de 80%. A Figura 6.4 ilustra o resultado desta mineração.



Figura 6.4. Resultado Mineração por Classificação feita pelo Diretor

Sua interpretação é a seguinte:

- Se o desempenho em “Prova” for “Excelente” então o aluno “Trancou” o curso;
- Se o desempenho em “Prova” for “Regular” então o aluno é “Desistente”;
- Se o desempenho em “Prova” for “Ótimo” e a região de procedência for “BA” então o aluno é “Desistente”.

### 6.3. Considerações Finais

Este capítulo apresentou um estudo de caso ilustrando o sistema Midas-Poeta em execução. Para realizar este estudo, utilizaram-se dados fictícios de um curso fictício de pós-graduação em macroeconomia à distância.

Demonstramos um cenário de execução apresentando as interações de um professor e do diretor do curso com o sistema. O professor utilizou a mineração por associação, enquanto que o diretor minerou regras de classificação.

Obviamente, considerando que os dados utilizados no processo de mineração são fictícios, não cabe aqui uma discussão sobre validade ou significado práticos do padrão de conhecimento produzido. Porém, o que se mostra interessante é que com esta abordagem de acompanhamento do aprendizado baseado em mineração de dados, aliado a concepção do modelo analítico de dados, o usuário terá a possibilidade de correlacionar dados do acompanhamento informal, do acompanhamento alternativo,

dados demográficos, comportamentais e dados sobre avaliação ao mesmo tempo e sem nenhuma regra ou restrição de seleção. Com isso, pode-se produzir informações gerenciais que o auxiliem na difícil tarefa de acompanhar o processo de ensino/aprendizagem em educação à distância.



### CONCLUSÃO

Quando planeja o ensino, o professor fixa os objetivos que pretende alcançar, que são os conhecimentos que o aluno deve adquirir ou formar durante o processo de aprendizagem. Sendo assim, é responsabilidade do professor utilizar os meios adequados para garantir o acompanhamento do aprendizado de forma satisfatória. Somente através deste acompanhamento é que o mesmo tem a possibilidade de diagnosticar problemas de ensino/aprendizagem e executar ações que favoreçam a construção do conhecimento por parte dos alunos.

No ensino presencial o professor pode utilizar-se de uma série de instrumentos de acompanhamento, tais como os testes elaborados pelo professor e as técnicas de observação em sala, proporcionando informações que garantem a verificação do aproveitamento, aptidões e ajustamento pessoal-social de seus alunos.

Aliada a esta base de informações, uma tendência atual no acompanhamento da aprendizagem é o processo de análise de dados, que correlaciona fatores como dados demográficos e comportamentais, dados perceptivos e dados sobre as avaliações dos alunos, agregando informações sobre a vida escolar do estudante. Estes dados podem ser úteis tanto para a análise individual de alunos ou grupos de alunos, como de problemas inerentes a uma disciplina sendo ensinada.

Na Educação a Distância – EAD, a avaliação e o acompanhamento do aprendizado são os principais problemas de sua implementação. As principais causas disso são a falta de contato corpo-a-corpo entre professores e alunos e a falta de uma prática pedagógica de acompanhamento voltada especificamente para EAD.

Por conseguinte, alguns trabalhos voltados para a implementação da EAD têm investigado como se pode realizar o acompanhamento de alunos em cursos à distância. Alguns deles se destinam ao acompanhamento informal, onde as interações do aluno com o ambiente de EAD são analisadas de forma a acompanhar o grau de motivação e participação, em analogia à observação de sala de aula no ensino presencial. Outros

estudos se destinam ao desenvolvimento de um acompanhamento alternativo aplicando avaliação autêntica e formativa em educação à distância.

Este trabalho apresentou uma estratégia que utiliza técnicas de mineração de dados para descobrir padrões sobre o processo de ensino/aprendizagem em EAD, auxiliando o professor em suas decisões pedagógicas. Seu ponto forte é a definição de um Modelo Analítico de Dados – MAD, que contém um conjunto de dados sobre o aluno e passíveis de análises.

Tal estratégia apresenta como principais características a simplicidade e a praticidade. Praticidade porque foi usada no desenvolvimento de um sistema real, e conseqüentemente, demonstra a sua aplicabilidade. A simplicidade é evidenciada dada a utilização de recursos e tecnologias existentes como *views*, banco de dados e SQL.

Utilizando a estratégia de acompanhamento, foi desenvolvido o Midas-Poeta, concebido como um módulo de análise de dados para o ambiente de EAD chamado Portfolio-Tutor.

A concepção do MAD, disponível no Midas-Poeta, foi criada a partir da base de dados do sistema Portfolio-Tutor e em sua implementação foi utilizado um conjunto de visões, cada uma refletindo uma informação de interesse.

O diferencial com relação aos ambientes existentes é que o Midas-Poeta agrega num único repositório os aspectos do acompanhamento informal e alternativo associados os dados demográficos, comportamentais e sobre avaliação, constituindo-se assim, num conjunto de dados relevantes para a tomada de decisão sobre o acompanhamento do aprendiz. Além do mais, o sistema permite que o usuário (professor ou diretor) realize uma consulta correlacionando os fatores que desejar, sem restrições de cruzamento de dados. Adicionalmente, por aplicar mineração de dados, o Midas-Poeta provê a identificação semi-automática de padrões que podem ser úteis ao usuário em sua tarefa de acompanhar o processo de ensino/aprendizagem.

Apesar de ter sido desenvolvido para atender aos professores e diretores, o Midas-Poeta poderia ser aplicado a um sistema Tutor. Nesta aplicação, o sistema Midas-Poeta se apresentaria para o sistema Tutor como um módulo de aquisição de conhecimento, permitindo que o comportamento do tutor pudesse se adaptar ao real desempenho de um grupo de alunos. Ou seja, a partir dos resultados da mineração, um

conhecimento em forma de regras poderia ser obtido, de forma que tais regras orientassem as estratégias de ação do tutor. A Figura 7.1 ilustra esta situação.



Figura 7.1. Integração do sistema Midas-Poeta com um sistema Tutor

## 7.1. Contribuições

- Descrição de uma estratégia para o acompanhamento do aprendizado em EAD, que permite a coleta de um conjunto de dados sobre o aluno e possibilita a análise desses dados para encontrar padrões que auxiliam o professor na melhoria do processo de ensino/aprendizagem;
- Incorporação de um recurso inovador – mineração de dados - para apoiar a tomada de decisões relacionadas ao acompanhamento do aprendizado;
- Desenvolvimento de um ambiente de descoberta de conhecimento aplicado à educação;
- Utilização de recursos na implementação do ambiente que possibilitam o acesso via Internet, permitindo a realização da mineração de dados remotamente;
- Definição de um conjunto-referência de informações para o acompanhamento do aprendizado em ambientes de EAD;
- Definição de um mecanismo que contempla o conjunto o conjunto-referência citado.

## 7.2. Trabalhos Futuros

A seguir, citamos alguns trabalhos que podem ser desenvolvidos como continuidade de nossas pesquisas:

- Integração das interfaces do Midas-Poeta com as interfaces do Portfolio-Tutor;

- Pesquisa, estudo e implementação de técnicas para a visualização gráfica dos padrões encontrados na mineração;
- Pesquisa e estudo de técnicas de escolha automática dos atributos para aplicação da mineração de dados. Um exemplo de pesquisa que concentra esforços neste sentido pode ser encontrado em Becker *et al.* (Becker, *et al.*, 2000);
- Integração do sistema Midas-Poeta com o Tutor-ACVA do Portfolio-Tutor. Estudos devem ser conduzidos no sentido de aplicar os resultados da mineração para orientar a execução do sistema tutor;
- Incrementar a estratégia de acompanhamento para apoiar a utilização de ferramentas de *Text Mining* (Feldman, 1999). Estudos devem ser realizados no sentido de definir que informações seriam adicionadas a concepção de um MAD e como estas podem ser devidamente recuperadas para que a utilização das ferramentas de *Text Mining* pudesse ser efetivamente aplicada;
- Realizar uma análise de desempenho do sistema proposto.

## Referências Bibliográficas

---

- (Adams, *et al.*, 2000) ADAMS, B., ALWIN, L., BARTER, M., GRAGALE, R. P., GEE, D. E., LACY, J., PORTSCHELLER, P., SMIT, R., DAWSON, L., *Using Data to Improve Schools: What's Working?*. American Association of School Administrators. EUA. 2000.
- (Adriaans & Zantigue, 1996) ADRIAANS, P., ZANTIGUE, D., *Data Mining*. Addison Wesley Logman. Edingourgh Gate, England, 1996.
- (Agrawal, *et al.*, 1993) AGRAWAL, R., IMIELINSKI, T., SWANI, A., *Mining Association Rules Between Sets of Items in Large Databases*. SIGMOD 5/93, 207-216, Washington, EUA, 1993.
- (Agrawal & Srikant, 1994) AGRAWAL, R. SRIKANT, R., *Fast Algorithms for Mining Associations Rules*. In: XX VLDB Conference. Santiago, Chile, 1994.
- (Alves, *et al.*, 2002) ALVES, R. M., ERRICO, L., MESQUITA, R. C. *Safes: Um servidor de Avaliações Online para Ensino via Web*. II Seminário Nacional de Tecnologia para EAD. Universidade Federal de Uberlândia. Uberlândia, 2002.
- (Archbald & Newman, 1989) ARCHBALD, D., NEWMAN, F. *Assessing Achievement: Toward the Development of a New Science in of Educational Testing*. SUNY Press. New York, EUA, 1989.
- (Armstrong & Anthes, 2001) ARMSTRONG, J., ANTHES, K., *How Data Can Help*. National School Boards Association. American School Boards Journal. EUA. 2001.
- (Azevedo, 2000) AZEVEDO, W., *Aquiforium Educacional*. Seminário Virtual Panorama Atual da Educação a Distância no Brasil. São Paulo, 2000.
- (Becker, *et al.*, 2000) BECKER, J. D., BERKHIN, P., FREEMAN, P., *Automating Exploratory Data Analysis for Efficient Data Mining*. In: VI ACM International Conference on Knowledge Discovery and Data Mining. Boston. EUA. 2000.
- (Bernhardt, 1999) BERNHARDT, V. L., *Designing and Using Databases for School Improvement*. Eye on Education. New York. 1999.
- (Bernhardt, 2000) BERNHARDT, V. L., *New Routes Open When One Type of Data Crosses Another*. Journal of Staff Development. Vol. 21, N. 1. EUA. 2000.
- (Bernhardt, 2001) BERNHARDT, V. L., *Intersections*. National Staff Development Council. Journal of Staff Development. Vol. 21. Nº 1. EUA. 2001.
- (Bloom, *et al.*, 1971) BLOOM, B. S., HASTINGS, J. T., MADAUS, G., *Handbook on Formative and Summative Evaluation of Student Learning*. McGraw Hill. Nova Iorque, EUA, 1971.
- (Booch, *et al.*, 2000) BOOCH, G., RUMBAUGH, J., JACOBSON, I., *UML – Guia do Usuário*. Editora Campus. 2000.
- (Boston, 2002) BOSTON, C., *The Concept of Formative Assessment*. Pratical Assessment, Research e Evaluation, ERIC Clearinghouse on Assessment and Evaluation. EUA. 2002.
- (Brumem, *et al.*, 2001) BRUMEN B.; WELZER T.; JAAKKOLA H., *Adaptive Incremental Framework for Performance Driven Data Mining*. Advances in Databases and Information Systems (ADBIS 2001), Lithuania, September 2001.

- (Cabena, *et al.*, 1998) CABENA, P., HADJINIAN, P., STADLER, R., VERHEES, J., ZANASI, A., *Discovering Data Mining From Concept to Implementation*. Prentice Hall PTR. New Jersey, EUA, 1998.
- (Cabral & Giraffa, 2000) CABRAL, A. R. Y., GIRAFFA, L. M. M., *WBT – Builder: Um Ambiente para Autoria de WBT*. In: XI Simpósio Brasileiro de Informática na Educação. Maceió. 2000.
- (Cai, *et al.*, 1991) CAI, Y., CERCONE, N., HAN, J., *Attribute-Oriented Induction in Relational Databases*. In: Knowledge Discovery in Databases. MIT Press, 1991.
- (Carter & Hamilton, 1998) CARTER, C., HAMILTON, H., *Efficient Attribute-Oriented Generalization for Knowledge Discovery for Large Databases*. IEEE Transaction Knowledge and Data Engineering. 1998.
- (Carvalho, 2000) CARVALHO, J. V., *Reconhecimento de Caracteres Manuscritos Utilizando Regras de Associação*. Dissertação de Mestrado. UFCG/COPIN. 2000.
- (Cendrowska & Stutz, 1987) CENDROWSKA, P., STUTZ, J., *PRISM: An Algorithm for Inducing Modular Rules*. International Journal of Man-Machine Studies. 1987.
- (Chen, *et al.*, 1996) CHEN, M., HAN, J., UY, P. S., *Data Mining: an Overview from Database Perspective*. In: Advances in Knowledge Discovery in Databases. Proceedings. 1996.
- (DBMiner, 2002) DBMiner Technology Inc: *DBMiner 2.0 (EnterPrise) User Manual*, 2002, 70 p.
- (Dietel, *et al.* 1991) DIETEL, R. J., HERMAN, J. L., KNUTH, R. A. *What Does Research Say About Assessment?* North Central Regional Educational Laboratory. Oak Brook. 1991.
- (Fayyad, 1998) FAYYAD, U. M., *Data Mining and Knowledge Discovery*. An International Journal. Kluwer Academic Publisher, V. 1, 1998.
- (Feldman, 1999) FELDMAN, R., *Mining Unstructured Data*. In: V ACM International Conference on Knowledge Discovery and Data Mining. San Diego. EUA. 1999.
- (Fields & Kolb, 2000) FIELDS, D. K., KOLB, M. A., *Desenvolvendo na Web com Java Server Pages*. Rio de Janeiro: Ciência Moderna, 2000, 549p.
- (Fisher, 1987) FISHER, D., *Improving Inference Through Conceptual Clustering*. In: AAAI Conference. Seattle, EUA, 1987.
- (Freitas, 2000) FREITAS, M. C. D., *As Mudanças na Relação Professor/Aluno Decorrente das Novas Tecnologias*. Laboratório de Sistemas de Apoio a Decisão. Universidade Federal de Santa Catarina. 2000.
- (Gotthard & Seiffert, 1997) GOTTHARD, A. M., SEIFFERT, R., *Mining Text Data*. DB2 Maganize, Vol. 2, N. 3. 1997.
- (Guedes, *et al.*, 2001) GUEDES, G. T. A., VICCARI, R. M., D'AMICO, C. B., *Uma Ferramenta para Auxiliar a Avaliação de Textos Construídos Colaborativamente em Ambientes de Ensino-Aprendizagem*. Cadernos de Informática, Vol. 2, N. 1. UFRGS, 2001.
- (Guttierres & Prieto, 1994) GUTTIERRES, F., PRIETO D., *A Mediação Pedagógica: Educação a Distância Alternativa*. Editora Papirus, Campinas, 1994.
- (Han, *et al.*, 1993) HAN, J., CAI, Y., CERCONE, N., *Data-driven Discovery of Quantitative Rules in Relational Databases*. IEEE Transaction Knowledge and Data Engineering. 1993.

- (Han, *et al.*, 1998) HAN, J., NISHIO, S., KAWANO, H., WANG, W., *Generalization-based Data Mining in Object-Oriented Databases Using an Object-Cube Model*. Data and Knowledge Engineering. 1998.
- (Han & Kamber, 2001) HAN, J., KAMBER, M., *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers. San Francisco. EUA. 2001.
- (Han & Fu, 1996) HAN, J., FU, Y., *Exploration of the Power of Attribute-Oriented Induction in Data Mining*. In: Advances in Knowledge Discovery in Data Mining. MIT Press, 1996.
- (Hernandez-Dominguez, 1998) HERNANDEZ-DOMINGUEZ, A., *Object-oriented Modeling of an Adaptable Virtual Class*. In: World Conferences on Educational Multimedia and Hypermedia on Educational Telecommunications. Alemanha, 1998.
- (Hernandez-Dominguez & Silva, 2000) HERNANDEZ-DOMINGUEZ, A., SILVA, A. S., *A Distributed Architecture of Intelligent Tutoring Systems in Distance Education Context*. In: Latin American Conference on Informatics, Cidade do México, 2000.
- (Hoffmann, 1993) HOFFMANN, J., *Avaliação Mediadora: Uma Prática em Construção da Pré-escola à Universidade*. Editora Mediação. 16ª Edição, Rio Grande do Sul, 1993
- (Holte, 1993) HOLTE, R.C., *Very Simple Classification Rules Perform Well on Most Commonly Used Databases*. Machine Learning. Vol. 11, pp. 63-91. 1993.
- (Huang, 1998) HUANG, Z., *Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values*. Data Mining and Knowledge Discovery, 2:283-304, 1998.
- (IBM, 2002) *IBM Intelligent Miner for Data*. Versão 6 Release 1, 2002, 410 p.
- (Jacobson, *et al.*, 1992) JACOBSON, I., CHRISTERSON, M., JONSSON, P., OVERGAARD, G., *Object-Oriented Software Engineering: A Use-Case Approach*. Addison-Wesley, USA, 1992.
- (Jaques & Oliveira, 1998) JAQUES, P. A., OLIVEIRA, F. M., *Agentes de Software para Análise das Interações em um Ambiente de Ensino a Distância*. In: Simpósio Brasileiro de Informática na Educação, 9., Ceará, Brasil, 1998.
- (Jain & Dubes, 1997) JAIN, A. K., DUBES, R. C., *Bootstrap Techniques for Error Estimation*. IEEE Transactions on Patterns Analysis and Machine Intelligence. 1997.
- (John & Langley, 1995) JOHN, G. H., LANGLEY, P., *Estimating Continuous Distribution in Bayesian Classifiers*. In: XI Conference on Uncertainty in Artificial Intelligence. San Mateo, EUA, 1995.
- (Johnson, 2000) JONSON, J. H., *Data-driven School Improvement*. ERIC Digest. ERIC Clearinghouse on Educational Management Eugene. Nº 109. EUA. 2000.
- (Kaufman & Rousseeuw, 1990) KAUFMAN, L., ROUSSEEUW, P. J., *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons. New York, EUA, 1990.
- (Kawasaki, 2000) KAWASAKI, E. I., *Um Modelo de Sistema de Tutoria Inteligente Baseado em Princípios Pedagógicos para a Educação de Adultos*. In: XI Simpósio Brasileiro de Informática na Educação. Maceió. 2000.
- (Keeney, 1998) KEENEY, L., *Using Data for School Improvement*. Annenberg Institute for School Reform at Brown University Providence. Houston. EUA, 1998.

- (Kinder, 2000) KINDER, A., *D3M: Helping Scholl Distill Data*. NCREL Learning Point. Vol. 2. Nº 2. EUA, 2000.
- (Kohavi, 1995) KOHAVI, R., *A Study of Cross-Validation and Botstrap for Accuracy Estimation and Model Selection*. Proceedings of the International Joint Conference on Artificial Intelligence. San Francisco, EUA, 1995.
- (Krasner & Pope, 1988) KRASNER, G. E., POPE, S. T., *A Cookbook for Using the Model-view-controller User Interface Paradigm in Smalltalk-88*. Journal of Object-Oriented Programming. EUA. 1988.
- (Larman, 2000) LARMAN, C., *Utilizando UML e Padrões: Uma Introdução à Análise e ao Projeto Orientado a Objetos*. Editora Bookman, PortoAlegre. 2000.
- (Lauritzen, 1995) LAURITZEN, S. L., *The EM Algorithm for Graphical Association Models with Missing Data*. Computational Statistics and Data Analysis, 19:191-201. 1995.
- (Lévy, 1999) LÉVY, P., *Cibercultura*. 34º Ed. São Paulo, 1999.
- (Linderman, 1986) LIDERMAN, R. H., *Medidas Educacionais*. Editora Globo. 1ª Edição. Rio Grande do Sul, 1986.
- (Luckesi, 2000) LUCKESI, C. C. *O que é Mesmo o Ato de Avaliar a Aprendizagem?* Pátio-Revista Pedagógica, Porto Alegre, ano 3. n. 12, p. 6-11, Fev/2000.
- (Machado & Becker, 2002) MACHADO, L. S., BECKER, K., *O Uso da Mineração de Dados na Web Aplicado a um Ambiente de Ensino a Distância*. In: I Workshop de Teses e Dissertações em Banco de Dados. XIX Simpósio Brasileiro de Banco de Dados. Gramado. 2002.
- (MacQueen, 1967) MACQUEEN, J., *Some Methods for Classification an Analysis of Multivariate Observations*. X Berkeley Symposium in Mathmatics, Statistics and Probability. 1:281-297, 1967.
- (Marietto & Omar, 2000) MARIETTO, M. G. B., OMAR, N., *Definição Dinâmica de Estratégias Instrucionais em Sistemas de Tutoria Inteligente: Uma Abordagem Multiagentes na Web*. In: XI Simpósio Brasileiro de Informática na Educação. Maceió. 2000.
- (Masetto, 2000) MASETTO, M. T. *Mediação Pedagógica e o Uso da Tecnologia*. In: Moran, J. M., Masetto, M. T., Behrens, M. A. *Novas Tecnologias e Mediação Pedagógica*. Editora Papirus, Campinas, 2000.
- (Melchior, 1999) MELCHIOR, M. C. *Avaliação Pedagógica – Função e Necessidade*. Editora Mercado Aberto. 2ª Edição. Rio Grande do Sul. 1999.
- (Menezes, et al., 1999) MENEZES, R. A., FUKS, H., GARCIA, A. B. *Utilizando Agentes no Suporte à Avaliação Informal no Ambiente de Instrução Baseada na Web – AulaNet*. In: X Simpósio Brasileiro de Informática na Educação. Fortaleza. 1999.
- (Nascimento, 2002) NASCIMENTO, D. M. C., *Um Sistema Tutor Acoplado a um Portfolio Eletrônico no Contexto da Educação a Distância – PortfolioTutor*. Dissertação de Mestrado, Universidade Federal de Campina Grande, 2002.
- (Neto, 1998) NETO, F. J. S. L., *EAD: Regulamentação, Condições de Êxito e Perspectivas*. Disponível em [www.intelecto.net/ead\\_textos/lobo1.htm](http://www.intelecto.net/ead_textos/lobo1.htm). Acesso em Out/2002.
- (Nunes, 2000) NUNES, I. B., *Noções de Educação a Distância*. Disponível em [www.intelecto.net/ead\\_textos/ivonio1.htm](http://www.intelecto.net/ead_textos/ivonio1.htm). Acesso em Out/2002.



- (Nurmela, *et al.*, 1999) NURMELA, K., LEHTINEN, E., PALONEN, T., *Evaluating CSCL Log Files by Social Network Analysis*. In: Computer Support for Collaborative Learning, Stanford, EUA, 1999.
- (Otsuka, *et al.*, 2002) OTSUKA, J. L., LACHI, R. L., FERREIRA, T. B., ROCHA, H. V., *Suporte a Avaliação Formativa no Ambiente de Educação a Distância TelEduc*. In: VI Congresso Iberoamericano de Informática Educativa. Espanha. 2002.
- (Protheroe, 2002) PROTHEROE, N., *Improving Teaching and Learning with Data-Based Decision: Asking the Right Questions and Acting on the Answers*. Educational Research Service. EUA. 2002.
- (Pyle, 1999) PYLE, D. *Data Preparation for Data Mining*. Morgan Kaufmann Publishers. San Francisco, EUA, 1999.
- (Quinlan, 1986) QUINLAN, J. R., *Induction of Decisions Tree*. Machine Learning, Vol. 1, Nº 1. pp 81-106. 1986.
- (Quinlan, 1993) QUINLAN, J. R., *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers. California, EUA, 1993.
- (Rahkila & Karjalainen, 1999) RAHKILA, M., KARJALAINEN, M., *Evaluation of Learning in Computer Based Education Using Log Systems*. 29<sup>th</sup> ASSE Frontiers in Education Conference. Porto Rico, 1999.
- (Ramos, 2000) RAMOS, R. Y., *Avaliar Conforme um Currículo Integrado com Temas Transversais*. Pátio – Revista Pedagógica. Porto Alegre, RS. Ano 3, Nº 12, 2000.
- (Resende & Barros, 2000) RESENDE, L., BARROS, L. A., *Auxiliando o Professor na Tarefa Metacognitiva de Estruturação do Conhecimento*. In: XI Simpósio Brasileiro de Informática na Educação. Maceió. 2000.
- (Richardson, 2000) RICHARDSON, J., *The Numbers Game: Measure Progress by Analyzing Data*. Tools for Schools. National Staff Development Council. EUA. 2000.
- (Rodrigues & Geyer, 2001) RODRIGUES, A. P., GEYER, C. F. R., *Agente Avaliação de Ensino e Aprendizagem em EAD*. Cadernos de Informática. Vol. 2, N. 1. UFRGS, 2001.
- (Santos, 1999) SANTOS, N. *Web-based Education: How to Assess Students Performance?* In: Site'99. Society for Information Technology and Teacher Education International Conference. USA. 1999.
- (Savasere, 1995) SAVASERE, A., *An Efficient Algorithms for Mining Association Rules in Large Databases*. In: XXI VLDB Conference, Zurich, Suíça, 1995.
- (See5, 1999) RULEQUEST RESEARCH, *Data Mining Tools See5 and C4.5*. Disponível em [www.rulequest.com/see5-info.html](http://www.rulequest.com/see5-info.html). December, 1999.
- (Seno & Vieira, 2002) SENO, W.P., VIEIRA, M.T.P. *DE-MAW: Um ambiente de Educação a Distância*. Revista Brasileira de Informática na Educação - RBIE, Volume 10, Número 1, pp. 21-29, abril de 2002.
- (Setzer, 2001) SETZER, V. W., *Os Meios Eletrônicos e a Educação: Uma Visão Alternativa*. São Paulo: Ed. Escrituras, Col. Ensaios Transversais, V. 10, 2001.
- (Silva, 2002) SILVA, D. R., *Um Ambiente para Descoberta de Conhecimento com Suporte de Data Warehousing e sua Aplicação para Acompanhamento do Aluno em Educação a Distância*. Dissertação de Mestrado. UFSCAR. 2002.

- (Silva & Fernandes, 2000) SILVA, J. C. T., FERNANDES, J. R., *AMON-AD: Um Agente Inteligente para Avaliação da Aprendizagem em Ambientes Baseados na Web*. In: VI Workshop de Informática na Escola. Curitiba. 2000.
- (Sistêlos, 1999) SISTÊLOS, A. J. C. M., *Um Ambiente Computacional de Apoio ao Método de Avaliação Autêntica: Projeto POETA (Portfolio Eletrônico Temporal e Ativo)*. Dissertação de Mestrado. UFPB/Copin. Campina Grande, 1999.
- (Souto, *et al.*, 2001) SOUTO, M. A., BICA, F., WARPECHOWSKI, M., VICARI, R. M., OLIVEIRA, J. P. M., ZANELLA, R., SONNTAG, A. A., *Ferramentas de Suporte a Monitoração do Aluno em um Ambiente Inteligente de Ensino na Web*. In: XII Simpósio Brasileiro de Informática na Educação. Vitória, 2001.
- (Souza, 1999) SOUZA, D. S., *Desafios da Gestão de Sistemas de EAD*. In: X Simpósio Brasileiro de Informática na Educação. Curitiba, 1999.
- (Sparks *apud* Schmoker, 2000) SPARKS, D. *Results Are the Reason*. National Staff Development Council. Journal of Staff Development. Vol. 21. Nº 1. EUA. 2000.
- (Srikant & Agrawal, 1996) SRIKANT, R., AGRAWAL, R., *Mining Quantitative Associations Rules in Large Relational Databases*. SIGMOD 6/96, 1-12, Montreal, Canadá, 1996.
- (Toebe, 2002) TOEBE, J. *Projeto e Construção de um Ambiente para Extrair Conhecimento de Bancos de Dados da Petrobrás. Estudo de Caso: Identificação Automática de Litofácies em Poços da Petrobrás*. Dissertação de Mestrado. UFCG/COPIN. Dezembro/2002.
- (Vieira, 2002) VIEIRA, M., *A Avaliação na Educação a Distância: Um Estudo sobre o Curso de Contemplação para Licenciatura em Biologia, Física, Matemática e Química*. II Seminário Nacional de Tecnologia para EAD. Universidade Federal de Uberlândia. Uberlândia, 2002.
- (Vrasidas & McIsaac, 1999) VRASIDAS, C., MCISAAC, M. S., *Factors Influencing Interaction in an Online Course*. The American Journal of Distance Education. Vol. 3, N. 3. 1999.
- (Wiggins, 1990) WIGGINS, G. *The Case for Authentic Assessment*. Practical Assessment, Research e Evaluation, ERIC Clearinghouse on Assessment and Evaluation. EUA. 1990
- (Witten & Frank, 2000) WITTEN, I. H., FRANK, E., *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann Publishers. San Francisco, EUA, 2000.

## TABELA DE DISCRETIZAÇÃO

VISÃO	FAIXA DE VALORES	VALOR DISCRETIZADO
Idade	Menor que 21	Jovem
Idade	21 .. 30	Adulto
Idade	31 .. 45	Maduro
Idade	Maior que 45	Velho
Interação p/ Material Didático	0	Nenhum
Interação p/ Material Didático	1 .. 3	Pouco Freqüente
Interação p/ Material Didático	4 .. 5	Regular
Interação p/ Material Didático	6 .. 7	Freqüente
Interação p/ Material Didático	Maior que 7	Muito Freqüente
Interação p/ E-mail	0	Nenhuma
Interação p/ E-mail	1 .. 10	Pouco Freqüente
Interação p/ E-mail	11 .. 20	Regular
Interação p/ E-mail	Maior que 20	Freqüente
Aproveitamento p/ Item	0	Zero
Aproveitamento p/ Item	0,1 .. 3,0	Péssimo
Aproveitamento p/ Item	3,1 .. 5,0	Fraco
Aproveitamento p/ Item	5,1 .. 6,9	Regular
Aproveitamento p/ Item	7,0 .. 8,0	Bom
Aproveitamento p/ Item	8,1 .. 9,0	Ótimo
Aproveitamento p/ Item	9,1 .. 10,0	Excelente
Aproveitamento p/ Artefato	0	Zero
Aproveitamento p/ Artefato	0,1 .. 3,0	Péssimo
Aproveitamento p/ Artefato	3,1 .. 5,0	Fraco
Aproveitamento p/ Artefato	5,1 .. 6,9	Regular
Aproveitamento p/ Artefato	7,0 .. 8,0	Bom
Aproveitamento p/ Artefato	8,1 .. 9,0	Ótimo
Aproveitamento p/ Artefato	9,1 .. 10,0	Excelente
Aproveitamento p/ Atividade	0	Zero
Aproveitamento p/ Atividade	0,1 .. 3,0	Péssimo

Aproveitamento p/ Atividade	3,1 .. 5,0	Fraco
Aproveitamento p/ Atividade	5,1 .. 6,9	Regular
Aproveitamento p/ Atividade	7,0 .. 8,0	Bom
Aproveitamento p/ Atividade	8,1 .. 9,0	Ótimo
Aproveitamento p/ Atividade	9,1 .. 10,0	Excelente