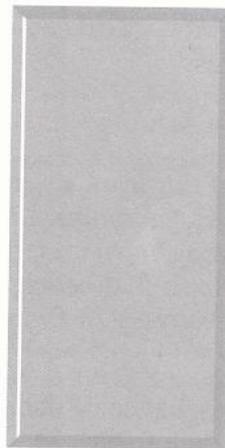




José Roberto Garcia (Uniso)
Juan Manoel Adan Coello (Puccamp)

*O ambiente dos negócios e a
extração de conhecimento de bases
de dados e a mineração de dados*



RESUMO

Este artigo tem como objetivo traçar um panorama sobre a evolução e as mudanças ocorridas nos ambientes dos negócios, a inteligência do negócio e o gerenciamento do conhecimento e relacionar estes acontecimentos a uma nova geração de técnicas e ferramentas pertencentes ao crescente campo de estudos da ECBD/MD – Extração de Conhecimento de Bases de Dados e a Mineração de Dados.

ABSTRACT

It is objective of this work to trace a scenario on: (a) the evolution and the changes the business environment has undergone, (b) the business intelligence and knowledge management and, (c) to relate these events to a new generation of techniques and tools that belong to the growing field of studies of the KDD/DM – Knowledge Discovery in Databases / Data Mining.

Palavras-chaves: banco de dados – gerência; negócios – processamento de dados; tecnologia da informação.

1. Histórico e evolução

Uma nova geração de técnicas e ferramentas está emergindo para assistir inteligentemente os seres humanos nas análises dos volumosos dados gerados diariamente pelas organizações e para encontrar conhecimento valioso e útil, em alguns casos, automaticamente. Estas técnicas e ferramentas pertencem ao crescente campo de estudos da Extração de Conhecimento de Bases de Dados/Mineração de Dados – ECBD/MD.

ECBD/MD é um termo que descreve uma variedade de atividades relacionadas a dados e informações. Também pode ser usado para descrever todo o processo visando descobrir padrões úteis nos dados analisados, incluindo não somente o passo de mineração de dados, que é executado através do uso de algoritmos específicos para descoberta de padrões, mas também o pré e o pós-processamento dos dados e padrões e uma grande quantidade de outras importantes atividades.

Assim, a Extração de Conhecimento de Bases de Dados é o processo de extrair informações válidas, previamente desconhecidas, compreensíveis e processáveis de grandes bases de dados visando a usá-las para tomar decisões cruciais no ambiente dos negócios (FAYYAD).

Embora a ECBD/MD se refira ao uso de métodos que podem automaticamente extrair informações de dados com pouca ou nenhuma intervenção de usuários, o termo, que ficou tão popular, agora pode referir-se a outros tipos de análises de dados como os baseados em consultas e geração de relatórios, processamento analítico *online* (OLAP) e análises estatísticas. Estes métodos não são estritamente sinônimos, porém são complementares em análises orientadas à descoberta e, na maioria das aplicações do mundo real, são usados cooperativamente. Também são usados para validar hipóteses prévias do conhecimento do negócio. Métodos analíticos orientados à descoberta são usados para identificar novos conhecimentos do negócio e para, automaticamente, refinar conhecimento prévio já validado.

Outro aspecto a ser considerado diz respeito à evolução dos dados do negócio para as informações do negócio, onde cada novo passo tem contribuído enormemente para essa evolução. Por exemplo, o acesso dinâmico aos dados é crítico para a recuperação do dado em aplicações de navegação dos dados, e a habilidade para armazenar grandes bases de dados é crítica para a mineração de dados. Do ponto de vista do usuário, os quatro passos

listados na Tabela 1.1, cujos dados referem-se ao mercado americano, foram revolucionários, pois permitiram que novas questões nos negócios fossem respondidas corretamente e com rapidez.

Tabela 1.1. Evolução da SI/TI¹ do ponto de vista do usuário.

Evolução	Questões dos Negócios	Tecnologias	Provedores dos Produtos	Características
Conjunto de dados (1960s)	“Qual foi o total de vendas nos últimos 5 anos?”	Computadores, fitas e discos	IBM, CDC	Entrega de informações retrospectivas e dados estáticos
Acesso aos dados (1980s)	“Quais foram as vendas unitárias na Bahia no último mês de março?”	Bases de dados Relacionais, Consultas com linguagem estruturada, ODBC	Oracle, Sybase, Informix, IBM, Microsoft	Entrega de informações retrospectivas e dados dinâmicos para o nível do registro
AD ² e Sistemas de Suporte à Decisão (1990s)	“Quais foram as vendas unitárias na Bahia no último mês de março? Forneça-me os valores de Salvador.”	Processamento Analítico <i>on-line</i> , Bases de Dados Multidimensionais, AD	Pilot, Comshare, Arbor, Cognos, Microstrategy	Entrega de informações retrospectivas e dados dinâmicos para múltiplos níveis
Mineração de Dados (ECBD/MD) (emergente)	“O que provavelmente acontecerá com as vendas unitárias em Salvador no próximo mês? Por que?”	Algoritmos avançados, Computadores multiprocessadores e base de dados massivas	Pilot, Lockheed, IBM, SGI, vários iniciantes (indústria nascente)	Entrega de conhecimentos prospectivos e proativos

¹ SI – Sistemas de Informação; TI – Tecnologia da Informação

² AD – Armazéns de Dados

Fonte: (PILOT)

2. O ambiente dos negócios

Mudanças fundamentais relacionadas ao cliente, ao produto e aos atos de comprar e vender, estão influenciando a forma das organizações verem e planejarem o relacionamento com os clientes. Com essas mudanças, surgem, conseqüentemente, requisitos que orientem uma nova forma de fazer negócios. Segundo CABENA, (a) as mudanças ocorridas nos ambientes dos negócios e (b) os requisitos necessários para acompanhar essas mudanças, são analisados a seguir.

(a) Mudanças ocorridas nos ambientes dos negócios◆ *Comportamento dos clientes:*

- Clientes mais exigentes e com acesso fácil e rápido a mais informações: guias de fornecedores, catálogos e Internet.
- Estrutura familiar alterada: somente 15% das famílias nos EUA mantêm estrutura familiar tradicional, ou seja, cônjuges, com ou sem filhos e em que somente o marido trabalha fora.
- Variedade de ofertas: muitos consumidores ficam confusos com a variedade de ofertas existentes e estão limitando o número de empresas com as quais estão preparados para negociar.
- Tempo para as compras: os consumidores estão valorizando mais o tempo gasto com compras de bens e serviços.

◆ *Saturação do Mercado:*

- Mercados saturados: nos EUA, a maioria das pessoas possui conta em banco, tem pelo menos um cartão de crédito, tem algum tipo de automóvel e seguro, e sabe bem onde comprar itens básicos de alimentação. Nessas áreas existem poucas opções para a expansão de mercado. As empresas normalmente baixam seus preços para poder competir, porém, a longo prazo essa estratégia não é boa.

◆ *Novos nichos de mercado:*

- Grupos étnicos: antes desvantajosos por possuírem conjuntos distintos de necessidades, hoje são vistos como o mercado consumidor urbano mais lucrativo onde as tendências de roupas e músicas há muito tempo estão emergindo. Como exemplo, a SunGlass Hut têm 1.600 lojas no mundo, vendendo, não somente um estilo ou modelo, mas centenas de diferentes estilos de óculos de sol, como, por exemplo, estilos dirigidos aos roqueiros e surfistas.

◆ *Aumento da comodidade:*

- Fornecedores estão desenvolvendo novas formas de distribuição e entrega de seus produtos; aumento de uso dos serviços *online*, desde

catálogos de seguros até malas diretas. As questões são: quem são os meus clientes e qual a melhor forma de alcançá-los?

◆ *Tempo para fazer negócios e o ciclo de vida curto dos produtos:*

- O tempo tem se tornado incrivelmente importante: as organizações necessitam surpreender os consumidores, cada vez mais rapidamente em função da concorrência, com produtos que lhes sejam agradáveis. A *Hewlett Packard* (HP) tem lançado uma nova impressora a cada seis meses; a *General Motors* (GM) lança no Brasil um novo modelo de carro a cada três meses; novos biscoitos são lançados no Brasil a cada quinze dias. A Netscape conseguiu obter aproximadamente 80% do mercado de navegadores com apenas um ano de estabelecimento.
- Computadores pessoais e Internet: as indústrias desses itens oferecem novos produtos e serviços mais rapidamente do que em qualquer outra época da história da computação. Lou Gerstner, Chairmain da IBM, recentemente cunhou a frase “*web year*” como sinônimo para três meses, quando se referia a essa velocidade.

◆ *Aumento da competição e negócios de risco:*

- As tendências para a comodidade, globalização, “*desregulação*” e a Internet aumentam as dificuldades em manter sob controle as forças de competitividade, as tradicionais e as novas.

(b) Requisitos necessários para fazer negócios

Muitas organizações têm sido forçadas a reavaliar as formas tradicionais de fazer negócios e têm procurado caminhos alternativos para responder às mudanças no ambiente dos negócios. Os principais requisitos e justificativas que orientam essa reavaliação estão descritos a seguir:

◆ *Foco no consumidor:*

1. Relacionamento: ênfase para a maior intimidade, colaboração e parceiro 1-para-1, isto é, possuir consumidores fiéis e exclusivos através da descoberta da necessidade do consumidor.

2. Questões novas estão surgindo, tais como:

- a) Quais classes de clientes eu tenho?
- b) Existem sub-populações com comportamentos similares aos clientes existentes e como endereçar mensagens comerciais para eles?
- c) Como vender mais para os clientes existentes?
- d) Quais entre os clientes existentes estariam propensos a comprar outros produtos ou serviços?
- e) Existem produtos ou serviços que meu cliente adquire no concorrente que eu poderia oferecer-lhe em curto espaço de tempo?
- f) Quais clientes provaram ser bons e fiéis e quais não? Existem clientes que dão “prejuízo”?
- g) É possível prever quais clientes têm um comportamento padrão nos pagamentos e quais são inconstantes podendo ser computados como risco?

◆ *Foco na Competição:*

Aumentar a força competitiva usando novas armas para o negócio. Algumas visões:

- a) Predição do potencial estratégico ou maior planejamento de negócios com os principais competidores: abertura de novas linhas de negócios ou territórios.
- b) Predição de movimentos táticos pelo local: abertura e localização de novos pontos comerciais, mudanças nos preços, oferecimento de novos serviços.
- c) Descobrir a existência de sub-populações de clientes que são especialmente vulneráveis aos concorrentes.

◆ *Foco sobre a vantagem do dado:*

Existe uma consciência crescente entre os administradores do negócio e os administradores de TI de que há atualmente grandes oportunidades decorrentes do uso da informação. Muitas organizações estão começando a ver os dados acumulados como recursos críticos para o negócio. Alguns dos fatores que contribuíram para essa consciência são:

- a) Crescente evidência do retorno sobre o investimento (ROI) exponencial de várias indústrias a partir de uma estratégia corporativa moder-

na para a tomada de decisões, baseada em técnicas orientadas aos dados, tais como AD.

- b) Crescente disponibilidade de AD. A adoção antecipada dessa nova tecnologia é uma forma de impulsionar a organização para a competitividade.
- c) Crescente disponibilidade de histórias de sucesso, publicadas pelos meios de comunicação apropriados.

3. A Inteligência do Negócio – IN (Business Intelligence) e o Gerenciamento do Conhecimento – GC (Knowledge Management)

A Inteligência do Negócio (IN) é um termo que tem sido usado para todos os processos, técnicas e ferramentas que apoiam as tomadas de decisão nos negócios, baseadas em TI. O Gerenciamento do Conhecimento (GC) é um termo que envolve comércio eletrônico, armazéns de dados, processamento analítico *online*, IN e ECBD/MD, entre outros, com o objetivo de melhorar o desempenho dos negócios através da adição, uso e controle do conhecimento descoberto.

A finalidade da IN parece simples: usar o conjunto de dados da organização para produzir melhores decisões nos negócios. Permitir que usuários finais e os analistas dos dados tenham acesso e analisem informações armazenadas em bases de dados transacionais, os dados de mercado e os armazéns de dados, são os objetivos pretendidos pelas ferramentas que estão sendo disponibilizadas diariamente. Segundo estima o Gartner Group, somente 3 a 5% dos dados corporativos são analisados pelas organizações (DATAQUEST). Porém, para atingir os objetivos citados, questões relacionadas à escalabilidade e acessibilidade, baseadas em arquitetura cliente-servidor e na *Web*, além da visibilidade, navegação e segurança dos dados, devem ser resolvidas.

Para atender ao crescente e vasto mercado da Inteligência do Negócio, grandes empresas têm investido vultuosas somas de dinheiro e enormes contingentes de profissionais desenvolvedores e outros especializados em *marketing*, além de estabelecer programas de parcerias com vendedores de ferramentas de análise. Assim, a IBM, uma das grandes empresas de computadores e *softwares* do mundo, anunciou o investimento de 70 bilhões de

dólares, a dedicação de um time composto por mais de 2.500 especialistas em IN e desenvolvedores, e um programa de parceria que já conta com 150 vendedores que se encarregarão de suprir, aos interessados, todo tipo de informação necessária sobre o assunto. Toda essa corrida tem uma explicação: o mercado de *software* para IN cresceu de US\$ 770 milhões em 1995 para US\$ 1.2 bilhão em 1997 (GREENGARD).

Consultores em IN têm fornecido inúmeros valores para provar a importância dessa tecnologia. Michael Haydock, vice-presidente para consultoria em inteligência nos negócios e serviços da *IBM's Global Consulting Group* diz que “durante os oito anos em que tem auxiliado organizações a implementar IN, o retorno sobre o investimento, após um ano, é de aproximadamente 10 para 1, e o retorno global sobre o custo de propriedade de uma solução IN gira em torno de 300%” (GREENGARD).

Segundo estimativa do International Data Corporation – IDC (IDC), as empresas constantes da *Fortune 500* perderiam 12 bilhões de dólares em 1999 devido à ineficiência no gerenciamento do conhecimento, resultando em retrabalho intelectual, desempenho abaixo do padrão e incapacidade para encontrar conhecimento. E o problema tende a piorar: estimativa para 2003 é que as empresas perderão o espantoso valor de 31.5 bilhões de dólares. Além disso, o gerenciamento do conhecimento, em termos mundiais, será o responsável pelo aumento do gasto de 2 bilhões, ocorrido em 1999, para mais de 12 bilhões de dólares em 2003.

Os componentes de IN

Segundo Michael Haydock (GREENGARD), quatro passos são necessários para uma implementação efetiva de IN: (1) coletar os dados, (2) minerar os dados, (3) decidir efetuar a transformação, e (4) gerenciar o conhecimento. Esses passos são fundamentalmente os descritos e utilizados no processo para a ECBD/MD. Essa situação é explicada porque IN e ECBD/MD pertencem ao vasto campo do gerenciamento do conhecimento, como citado anteriormente.

Além da metodologia que deve ser adotada, outro aspecto importante a ser considerado é o tempo de respostas. Com a rapidez com que acontecem as mudanças nos negócios nos dias de hoje, possuir a informação correta permite tomar decisões inteligentes rapidamente. Uma base de dados re-

lacional pode fornecer a velocidade e a facilidade de que o usuário necessita para lutar contra os competidores. As questões de negócios que o usuário precisa ter respondido envolvem, tipicamente, consultas complexas e processamento analítico *online*, índices dinâmicos e complexas uniões, isto é, complexos relacionamentos entre as informações, além de pesquisas em grandes bases de dados e obtenção de respostas rápidas para questões complicadas. Para atender a esses requisitos, os desenvolvedores estão disponibilizando ferramentas que possuam recursos como multiprocessadores simétricos, agrupamentos e sistemas de processamento paralelo massivo, tentando fornecer a confiabilidade e escalabilidade necessários. Também o gerenciamento do dado em ambiente paralelo e distribuído é importante.

Com o surgimento de tecnologias que permitem respostas rápidas e armazenamento de dados em bases multidimensionais possuindo índices múltiplos e dinâmicos, termos, agregados às funções que desempenham, estão surgindo no mercado de IN: *drill-down*, *drill-up*, *slice-and-dice*, *hypercubes*. As funções *drill-down / up* permitem ao usuário navegar, *online* e em tempo real, a partir de uma informação analítica, operacional (*drill-up*), atingir informações resumidas, e vice-versa (*drill-down*). *Slice-and-dice* permite “fatiar” a informação, segmentá-la por departamentos ou setores, ou mesmo nas mais diversas camadas – operacional, estratégica ou tática. Cubos multidimensionais – *hypercubes* – oferecem a facilidade de olhar a informação por vários ângulos, como por exemplo, as compras efetuadas pelo comprador “Antônio”, destinadas ao departamento de “produção” da fábrica do Rio Grande do Sul. Esta informação pode ser vista pelas dimensões do comprador, do departamento ou das unidades fabris, apenas com o pressionar de teclas.

Os desafios para IN

Raramente as empresas de médio e grande porte armazenam todos os dados departamentais e da empresa em um único gerenciador de bases de dados ou arquivo. Ao invés disso, dados-chave são gerenciados por múltiplos produtos (tais como *Oracle*, *Sybase*, *Informix*, *IMS*, *DB2*, etc.) que são executados em diferentes sistemas operacionais (*MVS*, *UNIX*, *Windows NT*) e, frequentemente, separados em várias localizações. Para sistemas IN, a heterogeneidade de muitos ambientes de computação apresenta vários desafios:

1. Prover aos usuários uma visão integrada dos negócios é a chave requerida para muitos sistemas de IN. Entretanto, *softwares* de IN devem ser flexíveis o suficiente para suportar diversas fontes de dados.
2. Similarmente, uma solução robusta de IN deveria suportar bases de dados heterogêneas. Muitas companhias estão adotando implementações distribuídas ou em múltiplas camadas, resultando em múltiplas bases de dados em uma única companhia. Com a adoção crescente da computação distribuída, cliente-servidor e departamental, é comum, para diferentes divisões de uma companhia, o uso de diferentes sistemas de base de dados.
3. A base de uma efetiva solução de IN é tornar fácil para o usuário final o acesso às informações relevantes do negócio. Um usuário pode necessitar consultar o armazém de dados da empresa para responder a uma questão, mas, para uma questão diferente, consultar uma base de dados local, ou para uma outra, a correlação dos dados cruzando-os com os dados de uma fonte externa. Tornar o usuário ciente da localização dos dados somado a complexidades desnecessárias é desencorajar o uso de todas as informações disponíveis.

Um sistema IN deveria proporcionar aos usuários as seguintes facilidades:

◆ **Acesso aos dados heterogêneos de forma transparente:** usuários dos negócios não precisam saber que eles estão conectados a múltiplas bases de dados. As interfaces deveriam providenciar leitura e gravação em diversas fontes de dados; deveriam gerenciar a complexidade de conexões para diferentes fontes; deveriam traduzir requisitos dos usuários para as interfaces daquelas fontes de dados, lidar com diferenças nos tipos de dados, e determinar uma estratégia eficiente de acesso ao dado para satisfazer aqueles requisitos dos usuários. Além disso, permitir que o dado particular, o dado externo, o armazém de dados da empresa, e o armazém de dados operacionais, sejam tratados como um único armazém de dados.

◆ **Consultas mais poderosas e mais simplificadas:** possuir habilidades para enviar uma única consulta para acessar dados armazenados em bases de dados relacionais ou não-relacionais, locais ou remotas permite construir aplicações poderosas para IN.

◆ **Acesso otimizado aos dados:** respostas rápidas para consultas são importantes para usuários de negócios. Mas muitas das ferramentas existentes

no mercado geram consultas pobres, resultando em baixos desempenhos e custos altos. O desafio é otimizar as consultas, com gravações automáticas e desempenho eficiente ocasionando custos mais baixos e resultados melhores.

♦ **Replicação heterogênea:** a capacidade para replicar dados é vital em muitas implementações de armazéns de dados. Permitir a atualização contínua das bases de dados com novos dados e facilitar, sempre, ao usuário do negócio o acesso ao dado mais atual, é importante para as aplicações de IN. Um outro ponto importante é a confiança que os dados devem apresentar aos usuários. Permitir capacidade avançada de replicação entre bases de dados heterogêneas, por exemplo, *DB2, Informix, Oracle, Sybase*, é outro ponto que aplicações de IN devem propiciar aos usuários dos negócios.

Reconhecer que soluções IN têm que operar em ambientes mistos e distribuídos de dados é o desafio dos ambientes de computação de hoje para cumprir as necessidades dos negócios do futuro.

4. Habilitadores que favorecem o uso de ECBD/MD

Existe um conjunto de habilitadores para a ECBD/MD os quais, quando combinados, motivam alterações substanciais nas formas tradicionais de tomada de decisão. Este conjunto de habilitadores é discutido a seguir:

♦ *Inundação de Dados:*

Quarenta anos de tecnologia de informação têm gerado enormes quantidades de dados (medidas em *gigabytes* e *terabytes*) sendo armazenados em sistemas de computadores. A computação, diariamente, tem acumulado dados sobre o comportamento das pessoas, e estes dados têm sido armazenados pelos bancos, companhias de cartões de crédito, sistemas de reservas e pontos eletrônicos de vendas. Um típico comércio de viagens gera informações sobre os hábitos dos viajantes e preferências em empresas aéreas, uso de cartões de crédito, material de leitura, serviços de telefones móveis, e talvez endereços na Internet. Dados demográficos e psicológicos têm sido uma fonte de dados úteis aos mineradores de informação. A disponibilidade de tais dados é particularmente importante visto que o foco da mineração de dados está sobre o comportamento do consumidor, o qual é freqüentemente orientado pelas preferências e escolhas que não são visíveis em uma simples base de dados.

◆ *Crescimento de AD:*

O crescimento dos Armazéns de Dados nas organizações tem fornecido a matéria prima básica para a mineração de dados: bases de dados limpos e bem documentados. A adoção antecipada da tecnologia dos armazéns de dados é fator de equilíbrio para a capitalização, mais adiante, de investimentos.

◆ *Novas soluções em TI:*

Soluções em TI para armazenagem e processamento têm possibilitado projetos para mineração de dados em larga escala. Isso é particularmente verdade com relação à tecnologia do processamento paralelo, pois muitos dos algoritmos para mineração são paralelos por natureza. Também os pacotes de ferramentas para visualização das informações fazem parte do arsenal para mineração de dados.

◆ *Novas pesquisas em aprendizado de máquina:*

Novos algoritmos vindos de centros de pesquisas e universitários estão pressionando os serviços comerciais mais rapidamente do que antes. A ênfase nas aplicações comerciais tem focado algoritmos melhores e mais escaláveis, vindo para o mercado através de produtos comerciais.

4.1. A evolução das tecnologias da informação e do conhecimento

A mudança do ambiente dos negócios é justificada pelo fato de que tomar decisões tem se tornado muito mais complicado – os problemas são mais complexos e os processos da tomada de decisões menos estruturados. Tomadores de decisão, hoje, necessitam de um conjunto de estratégias e ferramentas para orientar essas mudanças fundamentais.

Através da figura 1.1, pode-se observar o desenvolvimento ocorrido nos últimos 30 anos em TI, proporcionando informações melhores e mais refinadas.

Com estatísticas e relatórios, apenas informações resumidas foram disponibilizadas aos usuários. O dado poderia ser obtido pela requisição de um analista, isto é, a informação “resumo das vendas em um determinado período” teria que ser solicitada para um analista de dados.

Figura 1.1. Desenvolvimento da TI com relação ao refinamento da informação



Fonte: (DATAMINING)

Com os armazéns de dados, algumas consultas e relatórios podem ser feitos pelos próprios usuários dos negócios, por exemplo, relatórios de desempenho de produtos e lojas.

Com OLAP, resumos multidimensionais podem ser solicitados pelos usuários dos negócios, por exemplo, o total de vendas por produtos, por canal de distribuição, por mês.

Com o processo ECBD/MD, analistas de dados e um conjunto sofisticado de usuários dos negócios podem obter fatores influentes e tendências a partir dos dados. Frequentemente os analistas são necessários antes que as questões chaves sejam respondidas.

Com o acesso ao conhecimento, isto é, as informações obtidas através do processo de ECBD/MD, os usuários dos negócios podem utilizar as informações descobertas sempre que for necessário, embora novas informações sejam descobertas a cada semana ou a cada mês.

4.2. A situação da ECBD/MD

A ECBD/MD é um campo interdisciplinar que abrange técnicas como aprendizado de máquina, reconhecimento de padrões, estatística, bases de dados e visualização, com o objetivo de extrair conhecimento de grandes bases de dados. Esse campo de estudos surgiu a partir do momento em que as metodologias tradicionais de suporte à decisão, as quais combinam técnicas estatísticas simples com sistemas de informações executivas, não conseguem

fornecer a escalabilidade necessária para as grandes bases de dados e os armazéns de dados existentes hoje, bem como o tempo limite imposto pelo ambiente de negócios. A Mineração de Dados tem conseguido unir os mundos de negócio e o acadêmico, movendo-se, muito rapidamente, de um nicho de disciplinas de pesquisa, em meados dos anos 80, para um próspero campo, atualmente. De fato, 80% das empresas citadas em *Fortune 500* estavam, em 1997, envolvidas em um projeto piloto em mineração de dados ou já tinham desenvolvido um ou mais sistemas para mineração de dados (CABENA).

O interesse pela área de mineração de dados surge da confluência de duas importantes forças: a necessidade da descoberta de novas informações (conhecimentos), e os meios para tornar realidade a descoberta. A necessidade da descoberta decorre das mudanças ocorridas nos ambientes de negócios as quais têm resultado na crescente competitividade do mercado. Os meios referem-se principalmente aos avanços técnicos recentes em pesquisas com tecnologias em aprendizado de máquina, bases de dados, e visualização. Esta coincidência entre a pressão comercial crescente e os maiores avanços entre pesquisa e informação tecnológica emprestam um inevitável empurrão em direção à busca de conhecimento que poderá alavancar decisões nos negócios.

É incontestável, também, o desenvolvimento rápido que tem acontecido no campo da mineração de dados, abrangendo áreas distintas e atingindo toda a organização e todos os tipos de negócios. Apesar do desenvolvimento citado, alguns desafios persistem: a necessidade de descobrir formas para obter vantagem competitiva, através do uso do conhecimento resultante da análise dos grandes volumes de dados gerados diariamente e existentes nas corporações; a dificuldade de discernir o valor na informação – de separar o “bom” do “ruim” – e disponibilizar a “boa” informação para uso em toda a organização; a dificuldade de implementar ferramentas devido à inexistência ou ao desconhecimento de modelos que norteiem a implementação e uso dessas mesmas ferramentas; e a pouca informação existente no Brasil sobre o uso das técnicas e ferramentas para ECBD/MD.

5. Conclusão

A ECBD e a MD têm obtido bastante sucesso em curto espaço de tempo, e as expectativas geradas em relação a essa área têm crescido muito. Porém,

mesmo com o crescimento ocorrido nesses campos, ainda há carência de recursos apropriados que reduzam os riscos de falhas para os novos projetos. Por outro lado, se há um grande volume de informação publicada acerca da utilização efetiva e do potencial de utilização da ECBD em empresas dos EUA e Europa, o mesmo não acontece no Brasil, onde a bibliografia disponível a esse respeito é muito reduzida.

REFERÊNCIAS

- CABENA, P.; HADJINIAN, P.; STADLER, R.; VERHESS, J., ZANASI, A. *Discovering Data Mining: from concept to implementation*. New Jersey: Prentice-Hall PTR, 1998.
- DATAMINING – INFORMATION DISCOVERY INC. *The Knowledge Access Suite*. (on-line) 1999. (Citado em Abril 1999). Disponível na World Wide Web: <http://www.datamining.com/datamine/ka-suite.htm>
- DATAQUEST – ÍNDIA. *Intelligent Advantage*. (on-line) 1997. (Citado em jun. 1997). Disponível na World Wide Web: <http://www.dqindia.com/jun3097/3if2221101.html>
- FAYYAD, U.; UTHURUSAMY, R. Data Mining and Knowledge Discovery in Databases. *Communications of the ACM*, v. 39, n. 11, p. 24-26, nov. 1996.
- GREENGARD, S. *How to profit from Business Intelligence*. Beyond Computing (on-line) 1999. v. 8 n. 1 (Citado em julho 1999). Disponível na World Wide Web: <http://www.beyondcomputingmag.com/archive/1999/01-99/coverst.htm>
- IDC. *Fortune 500 companies will lose \$ 12 billion in 1999 due to Knowledge Management Inefficiencies*. (on-line) 2000. (Citado em jan. 2000). Disponível na World Wide Web: <http://www.idc.com/software/sw100599pr.htm>.
- PILOT SOFTWARE. *White Paper: an introduction to Data Mining*. (on-line) 1999. (Citado em abril 1999). Disponível na World Wide Web: <http://www.pilotsw.com/dmpaper/dmindex.htm>
- WEISS, S. M.; INDURKHIA, N. *Predictive Data Mining: a practical guide*. New York: Morgan Kaufmann Publishers Inc., 1998.

Endereço do autor:

Rua Arthur Caputti, 14
Jardim Guarujá
18050-040 – SOROCABA, SP
E-mail: jose.garcia@uniso.br