

## Padronização de Processos: BI e KDD

Nara Martini Bigolin

Departamento da Tecnologia da Informação -Universidade Federal de Santa Maria  
98400-000 – Frederico Westphalen – RS – Brazil

nara.bigolin@ufsm.br

***Abstract:** The Business Intelligence (BI) and the Knowledge Discovery in Databases(KDD) process are the extraction of interest patterns and features, which are not explicitly stored in databases. Such discovery may play an important role to understanding data, capturing intrinsic relationships between the information. This motivates the study and development of mechanisms to automates knowledge discovery for automatic knowledge acquisition. In this context this work presents a knowledge automatic acquisition cycle,integrating these two processes.*

***Key-Words:** Business Intelligence, Knowledge Management, Knowledge Discovery in Databases*

**Resumo:** Inteligência Empresarial e Descoberta de Conhecimento em Banco de Dados são os nomes dados a uma vasta categoria de programas aplicativos e tecnologias usadas para extrair, armazenar, analisar e transformar grandes volumes de dados, produzindo conhecimento capaz de auxiliar a empresa a tomar as melhores decisões nos negócios. Isso motivou o estudo e a aplicação de uma técnica de extração de conhecimento para aquisição automática de conhecimento. Nesse contexto, este trabalho propõe um ciclo de aquisição automática de conhecimento, integrando estes dois processos.

**Palavras-chave:** Inteligência Empresarial, Gestão de Conhecimento, Descoberta de Conhecimento.

### 1. Introdução

Tanto a inteligência empresarial (Davenport, 1998; Kimball 2010) quanto o processo de Descoberta de Conhecimento em Banco de Dados (DCBD) (Fayyad, 1996, Herbert 1998) utilizam técnicas de extração de conhecimento para obter conhecimento implícito. O DCBD é um conceito utilizado para denominar a exploração de informações implícitas em grandes volumes de dados. Essa tecnologia surgiu pela necessidade e pela dificuldade de explorar grandes bancos de dados de empresas. A inteligência empresarial é o conhecimento e a previsão dos ambientes interno e externo à empresa, orientando as ações gerenciais, tendo em vista a obtenção de vantagens competitivas (Choo, 1998). Neste contexto, dois enfoques diferentes, porém complementares, foram desenvolvidos, na tentativa de solucionar esses problemas: um, pela comunidade de Banco de Dados e outro, pela comunidade de Inteligência Artificial. A comunidade de Banco de Dados desenvolveu Sistema de Gerenciamento de Banco de Dados (SGBD), que tem por objetivo oferecer ferramentas que possibilitam o armazenamento e a manipulação de grande quantidade de informações estruturada e um

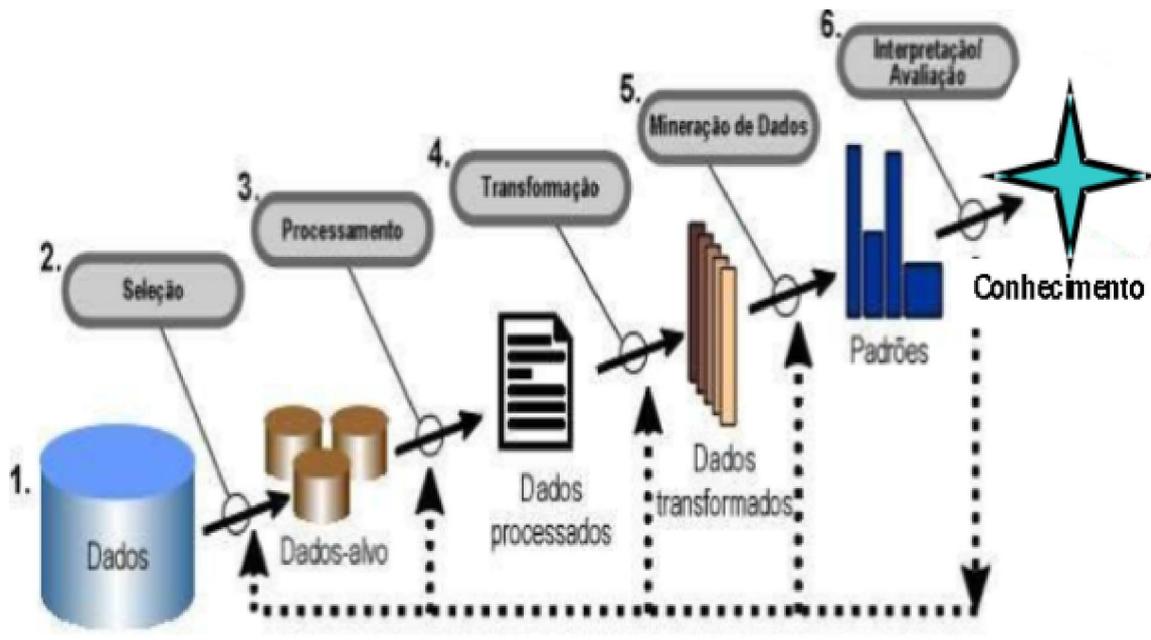
acesso rápido a elas. A maior preocupação no desenvolvimento deste tipo de sistema é o aspecto de modelagem dos dados, as linguagens de consulta e a eficiência na recuperação da informação. A comunidade de Inteligência Artificial interessou-se pela mineração de dados (extração de informações implícitas a partir de pequenas quantidades de dados) e a aprendizagem a partir de um pequeno conjunto de informações. A combinação dessas duas abordagens originou a área de Descoberta de Conhecimento em Banco de Dados. As técnicas tradicionais de aprendizagem (em inglês Machine Learning) descobrem conhecimentos implícitos a partir de dados extraídos do mundo real, enquanto as técnicas de mineração utilizam dados extraídos de bancos de dados. No primeiro caso, os dados de entrada para a mineração são representados em uma estrutura simples, e as informações relevantes são definidas pelo especialista. Já no segundo caso, além da estrutura e os tipos de dados serem complexos, o volume das informações é muito grande e, parte deste conjunto, pode ser irrelevante. Assim, é necessário efetuar dois tratamentos: um, para encontrar um subconjunto de informações apropriado (limpeza) e outro para adaptar essas informações em uma estrutura (transformação) aceitável pelas ferramentas de mineração de dados.

A seleção das informações relevantes é tão importante quanto o reconhecimento de padrões, pois sem os dados apropriados, dificilmente algum conhecimento útil será extraído do banco de dados. Para selecionar a amostra adequada é fundamental que haja um entendimento do domínio da aplicação e um prévio conhecimento dos dados relevantes, para que se possa estabelecer as metas do processo de descoberta de conhecimento implícito do ponto de vista do usuário. O segundo tratamento consiste em simplificar a estrutura complexa dos dados para um formato reconhecível pelas técnicas de mineração de dados.

Neste contexto, este artigo propõe um ciclo para aquisição automática de conhecimento baseado no processo de descoberta de conhecimento em banco de dados e no ciclo de inteligência empresarial. O restante do artigo está organizado da seguinte forma: na próxima seção são descritos o processo de descoberta de conhecimento e o ciclo de inteligência empresarial. Na terceira seção, a proposta do ciclo de aquisição automática de conhecimento é apresentada. Finalmente, as considerações finais são discutidas.

## **2. Processo de Descoberta de Conhecimento e Ciclo de Inteligência Empresarial**

Nesta seção serão descritos os dois processos que aquisição de conhecimento, suas etapas e funcionamento. Inicialmente, apresentaremos o processo de descoberta de conhecimento e, após, o ciclo de inteligência empresarial.



**Figura 1. Processo de Descoberta de Conhecimento**

O processo de DCBD (Fig. 1) (Fayyad, 1996) e o ciclo de inteligência empresarial (Fig. 2) (Herring 1997) têm o mesmo objetivo, auxiliar na gestão de informação, através de extração de informações implícitas em grandes bancos de dados. O processo de descoberta de conhecimento envolve cinco etapas sucessivas: seleção, processamento e transformação dos dados; mineração de dados; interpretação/avaliação do conhecimento. Essas etapas podem ser generalizadas em três grandes fases: pré-processamento, mineração e pós-processamento de dados.

O pré-processamento é uma das fases mais demoradas do processo de DCBD, pois consome cerca de 80 % dos esforços necessários para concluir todo o processo. Nessa fase são realizadas as tarefas de seleção, de processamento e de transformação dos dados. A seleção de dados consiste em selecionar os dados relevantes, para resolver um problema específico. Na etapa de processamento e tratamento de dados, os dados selecionados devem ser tratados e colocados em uma estrutura apropriada que será utilizada na etapa de mineração de dados.

A mineração de dados consiste em extrair modelos a partir dos dados selecionados e processados através do uso de algoritmos de aprendizagem. Esses algoritmos permitem a classificação e a caracterização de um conjunto de dados. O resultado desta etapa é um modelo de conhecimento, que pode ser na forma de grafos, de árvores de decisão ou de regras do tipo se premissa então conclusão (Han, 2011).

O pós-processamento é realizado através das etapas de avaliação e apresentação dos padrões, que são responsáveis pela identificação e análise dos padrões interessantes que representam informação implícita, bem como, a forma como esta informação será apresentada ao usuário. A interpretação/validação formata o conhecimento obtido e apresenta ao usuário ou reutiliza no sistema.

Para Herbert (1998) “inteligência empresarial é o conhecimento e previsão dos ambientes interno e externo à empresa, orientando as ações gerenciais, tendo em vista a obtenção de vantagens competitivas”.

A inteligência empresarial vem como uma tecnologia inovadora na implementação de soluções nesta área, sendo fundamentada em benchmarking e em pesquisa acadêmica. Benchmarking é o processo contínuo de comparar resultados de uma empresa, através de indicadores que irão ajudar a empresa a implementar ações que melhorem suas performances. A inteligência empresarial segue a linha do data warehouse evolutivo.

A inteligência empresarial segue três princípios básicos: modularidade, alta orientação ao negócio do cliente e trabalho cooperativo. O conceito de modularidade é utilizado em todos os níveis da inteligência empresarial. É a capacidade de dividir em módulos integráveis formadores do data warehouse evolutivo, possibilitando que sejam desenvolvidos em processos separados. A inteligência empresarial é orientada ao negócio do cliente, oferecendo aos tomadores de decisões uma visibilidade oportuna e consistente dos pontos - chaves dos seus negócios, alcançando todas as funções empresariais de sua empresa de forma integral, clara e objetiva. Para permitir uma melhor aplicação e concentração de recursos humanos e de tempo, o trabalho cooperativo vem com responsabilidade e requisitos bem definidos dos processos. O envolvimento do cliente é essencial no desenvolvimento de projetos de inteligência empresarial e a participação ativa do cliente sustenta a objetividade e a transparência da tecnologia de inteligência empresarial.

**A tecnologia da informação auxilia no sentido de possibilitar que a partir de dados, obtém-se informações, seguido de conhecimento para auxiliar nas ações a serem tomadas pela empresa.**



**Figura 2. Fluxo de informações**

Este fluxo de informação é obtido através do uso do ciclo de inteligência empresarial. Este ciclo, adaptado por Herring 1997 é apresentado na abaixo.

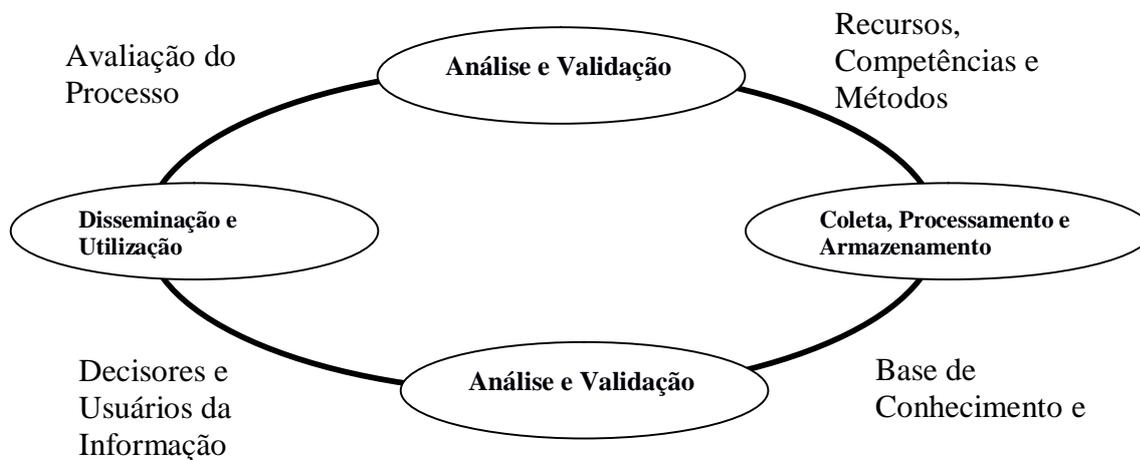


Figura 3. Ciclo da Inteligência Empresarial

O ciclo de inteligência empresarial (Terra 2000) é composto de quatro fases: planejamento e coordenação; coleta, processamento e armazenamento; análise e validação da informação; disseminação e utilização.

O **planejamento e coordenação** é a fase na qual são definidos: as necessidades de informatização da empresa, o banco de dados e as funcionalidades do sistema, assim como, os produtos que utilizarão a inteligência empresarial.

A **coleta, processamento e armazenamento da informação** envolvem a busca de fontes de informações para o atendimento às necessidades de estratégia do negócio.

Na **análise e validação da informação** as informações coletadas na etapa anterior, muitas delas sem uma aparente conexão entre si, são analisadas nesta etapa pelas redes de especialistas da área, com o objetivo de verificar a consistência das informações, estabelecer relações e avaliar o impacto destas para a organização.

A **disseminação e utilização da inteligência** encerram o ciclo de inteligência empresarial, quando disponibilizam o resultado do processo para os usuários da inteligência. As mudanças ocasionadas pela tecnologia usada para gerar, disseminar, acessar e usar a informação demanda por habilidades e competências relacionadas ao acesso, uso e disseminação da informação (Paletta, 2014).

Na seção seguinte, é proposto e descrito o ciclo de aquisição automática de conhecimento que foi baseado no processo de descoberta de conhecimento em banco de dados e no ciclo de inteligência empresarial.

### 3. Ciclo do processo de aquisição automática

O ciclo foi baseado no processo de descoberta de conhecimento proposto por Fayyad, (1996) e pelo ciclo de Inteligência Empresarial adaptado por Herring (1997). O objetivo final é alimentar um banco de dados com conhecimento adquirido do próprio banco de dados. O ciclo tem oito fases: seleção dos dados, limpeza dos dados, tratamento dos dados, extração do conhecimento, avaliação do conhecimento, interpretação do conhecimento, modelagem do conhecimento e inclusão no banco de dados.

O processo inicia pela **seleção dos dados**, a qual é realizada pelo auxílio da linguagem de consulta do Sistema Gerenciador do Banco de Dados vinculado à aplicação. Nesta etapa são selecionados os dados relevantes para fazer a extração posterior do conhecimento; a **limpeza dos dados** consiste em retirar os valores inválidos, incorretos e irrelevantes para a extração de conhecimento; a **transformação dos dados** tem o objetivo de formatar os dados numa estrutura compatível com as ferramentas de extração de conhecimento, que aceitam somente formato simples do tipo atributo/valor e com valores discretizados; a **extração** é responsável pela mineração de dados e extração dos padrões a partir dos dados fornecidos; a **avaliação de conhecimento** tem o objetivo de analisar o que foi gerado, se o mesmo é válido e relevante; na **interpretação**, o auxílio de um especialista do domínio é requerido, para classificar o conhecimento obtido em comprovado, novo ou irrelevante; uma vez o conhecimento validado, a **modelagem** deste conhecimento é efetuada para a inserção dentro do banco de dados. Nesta etapa uma metodologia de modelagem deve ser utilizada e, finalmente, na fase de **inserção**, o banco de dados é alimentado com novo conhecimento descoberto.

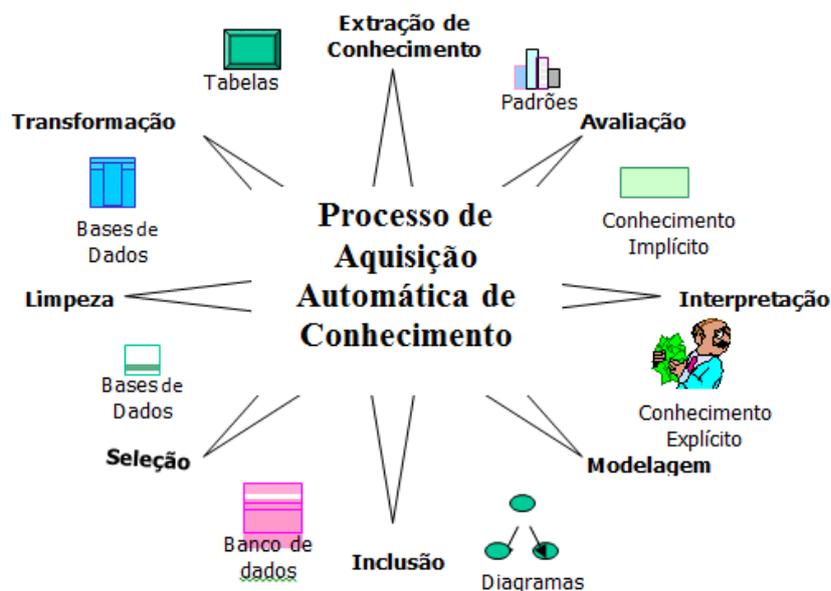


Figura 4. Ciclo da Aquisição Automática de Conhecimento

No contexto de uma análise comparativa entre as três metodologias (o processo de descoberta de conhecimento em banco de dados, o ciclo de inteligência empresarial e o ciclo de aquisição automática de conhecimento), pode-se afirmar que a última permite, devido ao nível de simplicidade e detalhamento, que usuários não especialistas de informática possam extrair conhecimento a partir de bancos de dados.

Abaixo, segue o quadro das fases do ciclo de AAC, do processo de DCBD e do ciclo de Inteligência Empresarial.

Tecnologias	Etapas			
Ciclo de AAC	Domínio da Aplicação	Seleção, Limpeza e Transformação	Extração e Avaliação do Conhecimento	<b>Modelagem e Inclusão no Banco de Dados</b>
Processo de DCBD		Seleção e Processamento Tratamento	Mineração Análise/ Interpretação	
Ciclo de Inteligência Empresarial	Planejamento e coordenação	Coleta, Processamento e Armazenamento	Análise e Validação da Informação	Disseminação e Utilização da Inteligência

**Tabela 1: Quadro comparative**

O processo de descoberta de conhecimento se mostra mais restrito, comparado aos dois ciclos. O ciclo de Inteligência Empresarial tem seu foco na organização dos processos de uma empresa, enquanto que o ciclo de aquisição automática de conhecimento se situa na exploração das informações utilizando as tecnologias de informações disponíveis.

#### 4. Conclusão

Neste trabalho foi proposto um ciclo de aquisição automática de conhecimento e validada com cinco estudos de caso em domínios bem distintos. Este ciclo possibilita uma melhoria na qualidade de prestação de serviços das empresas, independente de seu domínio de atuação. Comprovou-se também que em todos os estudos de caso obteve-se conhecimento novo, relevante e muito importante para a tomada de decisão.

Outra conclusão importante é que a partir dos testes realizados, o ciclo de aquisição de conhecimento permitiu extrair informações ocultas e extremamente importantes para a tomada de decisões, que com a utilização de uma ferramenta convencional e experiência de especialista não seria possível de obter-se.

A maior contribuição deste trabalho foi a proposta de um ciclo simples e capaz de ser usado em qualquer domínio de aplicação e por profissionais de qualquer área. Neste contexto, a aplicabilidade do ciclo de aquisição de conhecimento em quase todos os domínios de aplicação, torna esta tecnologia como a mais promissora do momento e com sucesso garantido para todas as empresas.

Como trabalhos futuros, se pretende modelar, através da metodologia UML, todos os conhecimentos adquiridos e inseri-los em um banco de dados, construindo assim um banco de conhecimento. Todo o conhecimento adquirido e não armazenado pode ser perdido. Uma vez armazenado num banco de dados, pode-se utilizar como fonte de novas descobertas de conhecimento. Pretende-se também integrar as quatro etapas no planejamento estratégico de uma empresa.

## Referências Bibliográficas

Choo, C. W. *The knowing organization: how organizations use information to construct meaning, create knowledge, and make decisions*. Oxford University Press, 1998.

Davenport T.H. and Prusak, L. *Conhecimento Empresarial: como as organizações gerenciam o seu capital intelectual*. 5 ed. Rio de Janeiro, Campus, 1998,

Fayyad, U. et al.. Knowledge discovery and data mining: towards a unifying framework. In: *Proceedings of the 2nd International Conference on Knowledge Discovery & Data Mining*. Editors: Simoudis, E.; Han, J.; Fayyad, U. August 1996.

Han J., Kamber M. and Pei J. *Data Mining: Concepts and Techniques*, 3<sup>rd</sup>ed. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann Publishers, July 2011.

Herbert, A. Edelstein. *Introduction to Data Mining and Knowledge Discovery*, 2<sup>o</sup> Edition, Two Crows Corporation, 1998.

Herring, J. P. *Producing CTI that meets management needs and expectations* SCIP Competitive Technology Intelligence Symposium. Boston 1997.

Kimball R. Ross M., Thornthwaite W. Mundy J. Becker B. *The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence*. 2010.

Paletta, C.F. and Maldonado E. P. *Inteligência Estratégica e Informação Perfil profissional na Era da WEB 3.0*. Revista Inteligência Competitiva. São Paulo, v. 4, n. 2, p. 1-10, abr./jun. 2014.

Terra. J C. *Gestão do conhecimento: o grande desafio empresarial*. São Paulo: Negócio Editora, 2000.