

Descoberta de Conhecimento em Banco de Dados: Um Estudo de Caso da Pesquisa Científica na Universidade Federal de Lavras

Cristhiane Xavier Figueiredo
Universidade Federal de Lavras
Lavras - MG - Brasil

Olinda Nogueira Paes Cardoso (Orientadora)
Universidade Federal de Lavras
Lavras - MG - Brasil

Junho de 2004

Resumo

Este trabalho mostra como é possível otimizar a atividade gerencial das Universidades, utilizando a Descoberta de Conhecimento em Banco de Dados (DCBD) para auxiliar na Gestão do Conhecimento. Através da aplicação dessa técnica à Pesquisa Científica da Universidade Federal de Lavras (Ufla), o presente trabalho trouxe importantes informações e uma visão integrada de toda produção científica, tecnológica e bibliográfica desta instituição, possibilitando uma melhoria na Gestão do Conhecimento nesta instituição.

Palavras-chave: Descoberta de Conhecimento em Banco de Dados, *Data Mining*, Gestão do Conhecimento, Gestão de Universidades

1 Introdução

Na Sociedade da Informação e do Conhecimento, as organizações estão buscando alguma vantagem sustentável que as diferencie das outras em seu ambiente de negócio. O conhecimento passa a ser reconhecido como um dos mais importantes recursos de uma companhia, tornando possíveis ações inteligentes, inovações e capacidade de continuamente criar produtos e serviços excelentes em termos de complexidade, flexibilidade e criatividade. Assim, a Gestão do Conhecimento emerge como uma área que foca em como as organizações podem entender o que elas conhecem, o que elas necessitam conhecer e como elas podem tirar o máximo proveito do conhecimento.

Ao longo do tempo, percebeu-se que a velocidade de

coleta de informações era muito maior que a velocidade de processamento ou análise delas. Num ambiente extremamente mutável, torna-se necessário a aplicação de técnicas e ferramentas que agilizem o processo de extração de informações relevantes em grandes volumes de dados. A metodologia de Descoberta de Conhecimento em Banco de Dados veio preencher essa lacuna na necessidade de análise que ultrapassa habilidade e capacidade humanas.

Levando em consideração que, assim como as empresas, as Universidades precisam inovar, otimizando a atividade gerencial, principalmente, para acompanhar as exigências do mercado, e também vencer algumas dificuldades, como o gerenciamento dos dados, a visão extremamente segmentada, setorializada que a maioria tem delas mesmas e as diversas limitações encontradas na gestão dos sistemas de informações, o presente trabalho utilizou o processo de Descoberta de Conhecimento em Banco de Dados para extrair informações importantes que servem de apoio à tomada de decisão e à Gestão do Conhecimento na Universidade Federal de Lavras (Ufla).

2 Gestão do Conhecimento

A nova economia baseia-se em informação; o conhecimento e as competências essenciais são ativos organizacionais chaves. Produzir produtos ou serviços únicos ou produzi-los a um custo menor do que os concorrentes depende de um conhecimento superior sobre o processo de produção e de um projeto superior.

Saber como fazer coisas de forma eficaz e eficiente e

de modo que as outras organizações não possam copiar é uma fonte principal de lucro. Alguns teóricos da administração acreditam que esses bens de conhecimento são tão ou até mais importantes que os bens físicos e financeiros na garantia da sobrevivência e competitividade da empresa, afirma Laudon & Laudon (1999), ao citar Favela (1997).

Desenvolver procedimentos e rotinas para otimizar a criação, o fluxo, o aprendizado e o compartilhamento de conhecimento e informação numa empresa torna-se uma responsabilidade gerencial central. O processo de ativar e sistematicamente gerenciar e alavancar o armazenamento de conhecimento numa organização é chamado de Gestão do Conhecimento [Laudon & Laudon (1999)].

A Gestão do Conhecimento pode ser vista, então, como o conjunto de atividades que busca desenvolver e controlar todo tipo de conhecimento em uma organização, visando à utilização na consecução de seus objetivos. Este conjunto de atividades deve ter como principal meta o apoio ao processo decisório em todos os níveis. Para isso, é preciso estabelecer políticas, procedimentos e tecnologias que sejam capazes de coletar, distribuir e utilizar efetivamente o conhecimento, bem como representar fator de mudança no comportamento organizacional [Tarapanoff (2001)].

3 Descoberta de Conhecimento em Banco de Dados

O sucesso das organizações depende basicamente das decisões tomadas por seus gestores antes mesmo de apresentar ao mercado seus produtos ou serviços. Essas decisões têm se tornado necessárias em prazos cada vez mais curtos, exigindo dos tomadores de decisão uma atenção redobrada aos ambientes interno e externo da empresa. Muitas vezes, más decisões são tomadas, não pela inexistência do conhecimento para se tomar decisões melhores, e sim porque o conhecimento não estava disponível no tempo e lugares certos para serem utilizados.

Para que o conhecimento seja extraído de forma eficiente, é realizado um processo chamado Descoberta de Conhecimento em Banco de Dados (DCBD ou KDD - Knowledge Discovery in Databases), processo este que possui o *Data Mining* como principal etapa.

De acordo com Adriaans & Zantinge (1996), existe uma confusão entre os termos *Data Mining* e Descoberta de Conhecimento em Banco de Dados. O termo DCBD é empregado para descrever o processo de extração de conhecimento de um conjunto de dados. Neste contexto, conhecimento significa relações e padrões entre os elementos dos conjuntos de dados. O termo *Data Mining*, segundo os autores, deve ser usado exclusivamente para o estágio de descoberta do processo de DCBD. Este processo se divide em sete estágios: (1) Definição do Problema, (2) Seleção dos Dados, (3) Eliminação de incongruências/erros dos dados (filtragem dos dados), (4) Enriquecimento dos dados, (5) Codificação dos dados, (6) *Data Mining* e (7) Relatórios.

Segundo Fayyad et al. (1996), o termo DCBD refere-se a todo o processo de descoberta de conhecimento útil de um conjunto de dados e já o termo *Data Mining* refere-se à aplicação de algoritmos¹ para a extração de padrões em um conjunto de dados, mas sem os passos adicionais em um processo de descobrimento do conhecimento.

O processo de DCBD é feito em etapas, que envolvem a preparação dos dados, procura de padrões, teste do conhecimento e refino do modelo. É caracterizado por ser não trivial, ou seja, por possuir um grau de autonomia na procura pelo conhecimento.

São quatro as principais tarefas nas quais o analista se envolve: (1) Seleção ou Filtragem dos Dados, (2) Pré-processamento e Análise de Dados, (3) Transformação dos dados pela escolha do modelo de *Data Mining* e evolução e (4) Geração e Interpretação de Resultados. Em seguida, cada uma dessas etapas será observada com maior detalhe. A Figura 1 representa estas etapas do processo de DCBD.

4 Data Mining

Segundo Oracle (2001), no grande mercado competitivo, um dos fatores críticos para as empresas é o gerenciamento dos seus bens mais valiosos - seus clientes e as informações que elas têm sobre esses clientes. E é exatamente dentro deste contexto que o *Data Mining* pode ajudar. O *Data Mining* pode analisar minuciosamente grandes quantidades de dados e encontrar informações ocultas

¹ Conceito básico: algoritmo é uma seqüência de passos destinados a resolver um problema ou atingir algum objetivo específico

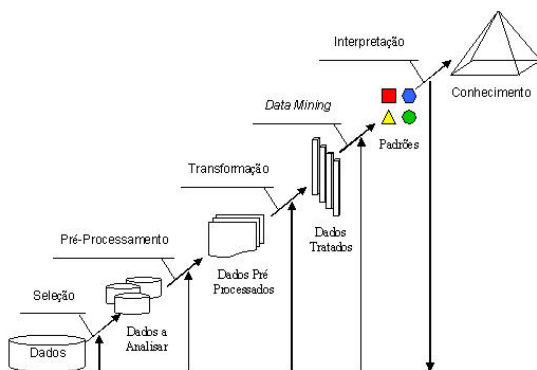


Figura 1: Processo de Descoberta de Conhecimento em Banco de Dados

Fonte: Santos, M. F. Descoberta de Conhecimento em Bases de Dados.

que podem ser vitais para o negócio da empresa, como, por exemplo, entender o comportamento dos clientes e atuar de acordo com as peculiaridades de cada um.

Segundo Davis King (2004), *Data Mining* é um modo de procurar relações interessantes escondidas em um grande conjunto de dados. Por interessante, pode-se entender qualquer estrutura necessária para investigação, como padrões de clustering e aproximações de funções. As técnicas de *Data Mining* podem ser a princípio semelhantes às análises de regressão. O método científico de *Data Mining*, consiste basicamente de cinco etapas: (1) Definir o problema, (2) Gerar hipóteses / modelos, (3) Coletar dados / conduzir experimentos que gerem dados, (4) Testar modelos em confrontação com os dados e (5) Utilizar os resultados para gerar novas hipóteses. Raramente é um processo completamente automatizado, com uma grande intervenção do analista que conduz o estudo.

Robert Grossman (2004) define *Data Mining* como a descoberta de padrões, associações, mudanças, anomalias e estruturas estatísticas e eventos em dados. A análise de dados tradicional é baseada na suposição, em que uma hipótese é formada e validada através dos dados. Por outro lado, as técnicas de *Data Mining* são baseadas na descoberta na medida que os padrões são automaticamente extraídos do conjunto de dados.

Em outras palavras, o *Data Mining* nada mais é do que um algoritmo matemático utilizado para produzir conhe-

cimento analisando dados, descobrindo tendências e fraudes, e assim, ajudar o usuário a chegar a conclusões que vão além da análise humana. *Data Mining* refere-se à garimpagem ou descoberta de novas informações em termos de padrões e regras oriundas de grandes quantidades de dados (de um *Data Warehouse*², por exemplo).

5 Plataforma Lattes

A plataforma lattes é um conjunto de Sistemas de Informações, bases de dados e portais Web voltados para a gestão de Ciência e Tecnologia (C&T). Foi concebida para integrar os sistemas de informações das agências federais, racionalizando o processo de gestão de C& T. Lançada em 16 de agosto de 1999 com a primeira versão do Sistema CV-Lattes, proporcionou um aumento significativo do número de currículos enviados ao CNPq, que chegou a mais de 100 novo currículos Lattes ao dia. Resultado do esforço conjunto do MCT, CNPq, FINEP e CAPES/MEC, a Plataforma Lattes constitui um importante passo para a integração dos sistemas de informação das principais agências de fomento do País, antiga demanda da comunidade científica e tecnológica.

A Plataforma Lattes integra, atualmente, quatro sistemas: O primeiro deles se refere a um Sistema Eletrônico de Currículos, que registra a vida pregressa e atual dos pesquisadores. O segundo sistema é o Diretório dos Grupos de Pesquisa no Brasil, uma base de dados que registra todos os grupos de pesquisa em atividade no País. O terceiro sistema é o Diretório de Instituições, instituições estas que demandam fomento ao CNPq. E, finalmente, o quarto sistema chama-se Sistema Gerencial de Fomento, cujo objetivo é dar possibilitar uma gestão estratégica para dar mais qualidade às atividades de fomento do CNPq.

Lattes Extrator

O Lattes Extrator é o instrumento de extração das informações disponibilizadas na Plataforma Lattes. Inicialmente está sendo disponibilizada a extração dos currículos Lattes e posteriormente das demais unidades de análise da Plataforma. Atualmente, as instituições licenciadas podem extrair, diretamente do banco de currículos

²*Data Warehouse* é um tipo de banco de dados que armazena uma quantidade muito grande de informações, que geralmente são utilizadas em sistemas de apoio a gestão.

Lattes do CNPq, os dados curriculares de seus pesquisadores, professores, alunos e colaboradores. O Lattes Extrator está limitado a extrair do banco de dados do CNPq os currículos de interesse da instituição, através de arquivos no formato XML. Com isto, as instituições podem criar seu próprio banco de currículos Lattes e, para tal, podem contar com o modelo e dicionário de dados do Currículo Lattes, disponibilizado pelo CNPq.

6 Metodologia

Após a coleta de informações necessárias para o embasamento teórico, foi realizado um estudo documental na Plataforma Lattes. Através de uma interface online do Lattes Extrator, como mostra a Figura 2, foram extraídos 575 currículos de professores, alunos, ex-alunos, mestrandos e doutorandos. Esses currículos estavam disponíveis como documentos no formato XML, o que implicou o desenvolvimento de um sistema para importar os dados desses documentos para um Banco de Dados. Antes disso, foi modelado e criado um Banco de Dados Relacional, contendo 58 tabelas, entre elas: Dados Gerais - com endereço profissional, formação acadêmica e atuações profissionais; Produção Bibliográfica, que incluía, entre outras, publicação de artigos e trabalho em eventos; Produção Técnica; Outra Produção e Dados Complementares, com Participação em Banca, Orientações Concluídas, entre outras.

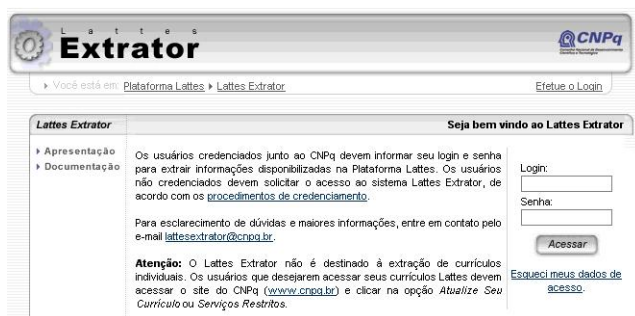


Figura 2: Interface Web do Lattes Extrator.

Implementado e povoado o Banco de Dados com mais de 18 mil linhas de dados, passou-se à etapa de filtragem dos dados, quando foram removidos dados inconsistentes,

campos em branco, dados repetidos, entre outros. Após a filtragem, iniciou-se a etapa de mineração dos dados ou *Data Mining*. Esta etapa foi subdividida em duas fases: a primeira consistindo basicamente de cruzamento de consultas SQL³ e a segunda, execução de algoritmos. Neste trabalho, apenas a primeira fase foi desenvolvida. Foram feitas várias consultas específicas para conhecer os dados inseridos e gerar alguns relatórios na forma de gráficos, permitindo análise e interpretação, resultando em conhecimento sobre a Pesquisa Científica na Ufla.

Dentre as etapas pré-definidas da técnica de DCBC, foram realizadas:

1. **Definição do Problema:** O maior problema identificado no decorrer do desenvolvimento deste trabalho foi a ausência de uma visão integrada dos processos de pesquisa científica da Ufla, impossibilitando o estabelecimento de regras, metas e estratégias de ação. A Ufla tem um alto potencial de desenvolvimento tecnológico que poderia ser melhor explorado, ou até mesmo diversificado, se fossem conhecidos padrões de procedimentos, associações de produções científicas, áreas do conhecimento com maior número de pesquisas, entre outras coisas.
2. **Seleção dos Dados:** Através do Lattes Extrator, foram selecionados e extraídos, inicialmente, documentos da Plataforma Lattes que continham os registros de toda produção científica de cada docente e de alguns alunos, ex-alunos, mestrandos e doutorandos da Ufla. Em seguida, foram selecionados, dentro destes documentos, dados específicos referentes às produções científica, tecnológica e bibliográfica dos mesmos, principalmente dos professores.
3. **Eliminação de incongruências/erros dos dados (filtragem dos dados):** Os dados selecionados na etapa anterior ainda continham algumas inconsistências, como, por exemplo, ausência de especificação de campos importantes e duplicação de outras especificações. Filtrando-se essas informações, passou-se às etapas seguintes.
4. **Enriquecimento dos dados:** Esta etapa do processo de DCBD não foi necessária na realização deste projeto, nenhuma informação teve de ser adicionada,

³Structured Query Language - linguagem de Banco de Dados.

visto que os documentos extraídos da plataforma latentes, na etapa 2, já continham todos os dados necessários à descoberta de conhecimento.

5. Codificação dos dados: Foram feitos dois tipos de codificação de dados. O primeiro consistiu da transformação dos documentos XML em documentos SQL, contendo o código de inserção dos dados e os dados a serem inseridos no Banco de Dados. O segundo tipo foi, basicamente, a execução desses códigos SQL, gerados na codificação anterior, no SGBD da Oracle.

Para a primeira codificação, fez-se necessário o desenvolvimento de um sistema que extraísse as informações contidas nos documentos XML e as inserisse em um Banco de Dados. Para isso, foram utilizados dois sistemas independentes: um software conversor de documentos XML em outros formatos - o Xalan C++, versão 0.40.0 - e o Sistema Gerenciador de Banco de Dados da Oracle - Oracle 9i DataBase.

Antes de iniciar a descrição do desenvolvimento dos sistemas, é necessário abordar os conceitos de XML e XSL. A *Extensibel Markup Language* (XML) é uma linguagem de marcação semelhante à HTML, porém, mais flexível no que se refere às marcas (tags) que ela utiliza. A XML permite a criação de novas tags, gerando uma estrutura totalmente definida para o documento. Cada dado ou informação possui uma descrição (metadado) e/ou um valor associado. Isso faz com que os documentos no formato XML se tornem legíveis para as pessoas e manipuláveis por computadores, permitindo, entre outras coisas, a recuperação de informações nesses documentos.

Apesar de todas essas vantagens, para a aplicação da Descoberta de Conhecimento em Banco de Dados sobre os dados coletados, foi necessário abrir mão das funcionalidades da XML para que isso fosse possível. Utilizou-se a linguagem XSL para identificar a estrutura dos documentos XML e depois, então, extrair desses documentos as informações necessárias a serem inseridas no Banco de Dados. A *Extensible Stylesheet Language* (XSL) é uma linguagem que permite criar folhas de estilo para documentos XML, permitindo, através da XSLT (*Extensible Styleshett Language Transform*), transfor-

mar documentos XML em diversos outros formatos, como HTML, texto simples ou qualquer outro documento baseado em texto, como, no caso deste trabalho, em um código SQL.

O documento XML possui uma estrutura hierárquica, podendo ser representado por uma estrutura de árvore. Essa característica permite que a folha de estilo, criada utilizando a XSL, identifique cada nó da "árvore"XML, ou seja, cada itag do documento XML, extraíndo seus valores e estruturando-os num código SQL que, depois, seria executado no Banco de Dados.

6. *Data Mining*: A etapa consistiu, basicamente, da primeira fase do *Data Mining*, com elaboração, cruzamento e comparação de consultas SQL. Foram definidas algumas consultas a serem feitas no Banco de Dados para conhecer os dados e identificar padrões, associações e regras que constituiriam a implementação da fase 2 do *Data Mining*.
7. Relatórios: Não foram gerados relatórios impressos que pudessem ser anexados a este trabalho. Resultados foram exibidos várias vezes e de várias maneiras a partir das consultas feitas na etapa anterior, permitindo a elaboração de um resumo dos principais pontos e a identificação dos elementos da ferramenta de *Data Mining* que será construída. Este resumo será apresentado na seção de Resultados e Discussões.

7 Resultados e Discussões

O pressuposto inicial de que há uma grande quantidade de informações e conhecimento "escondidos" nos registros da pesquisa científica da Ufla é bastante válido, apesar de ter sido verificado que os dados presentes nos currículos extraídos da Plataforma não estavam atualizados, o que representou uma limitação para este trabalho. Informações disponíveis no site oficial do CNPq (até o momento da realização deste trabalho), afirmam que a versão do Latentes Extrator que está disponível extrai apenas currículos atualizados até julho de 2002. Afirmam também que está sendo desenvolvida uma nova versão que permitirá a extração de currículos mais atualizados logo que estiver disponível.

Além desta limitação, outro fato importante deve ser levado em conta ao se analisar os resultados gerados por este trabalho: poucas pessoas atualizam seus currículos Lattes periodicamente. Um resultado crítico que advém desse fato é que dos 575 currículos inseridos no Banco de Dados, 90% não contêm atividades cadastradas. Essas atividades incluem atividades de ensino, pesquisa, direção e extensão, além de serviços técnicos e treinamentos ministrados. Os 10% das pessoas que incluíram suas atuações profissionais em seus currículos têm entre 2 e 61 atuações, demonstrando uma variedade muito grande de número de atividades. O gráfico da Figura 3 demonstra essa variedade.

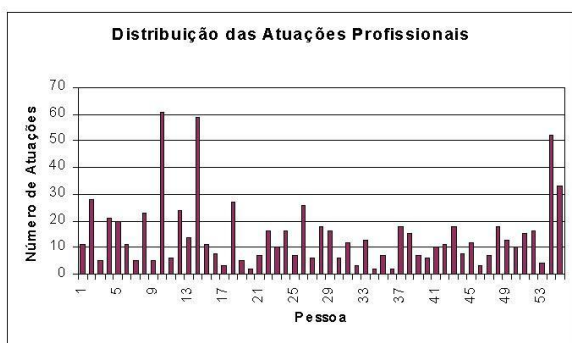


Figura 3: Número de atuações profissionais por pessoa.

O que se pôde observar das atividades cadastradas pelos 10% das pessoas, é que 6% delas realizaram entre 1 e 16 atividades de ensino. Essas atividades de ensino incluíam ensino médio, graduação, pós-graduação e especialização.

Nem todas as atividades cadastradas foram realizadas na Ufla, como algumas atividades de direção que incluíam gerência de empresas, por exemplo. Entre as atividades de direção, poderiam estar: coordenação de curso, chefia e subchefia de departamento, coordenação de laboratório, etc. Apenas 3% das pessoas cadastraram alguma atividade de direção, sendo que destas havia uma variação de 1 a 19 atividades por pessoa.

No que se refere às atividades de pesquisa, 7% realizaram entre 1 e 6 atividades de pesquisa. Cada uma dessas atividades compreendia, na maioria das vezes, de 1 a 3 linhas de pesquisa, mas algumas possuíam até 9 linhas. As atividades de extensão compreendiam 2,6% dos

cadastros, sendo que cada pessoa realizou entre 1 e 6 atividades de extensão. Além dessas, 2,4% realizaram entre 1 e 32 serviços técnicos e menos de 2% ministraram algum treinamento. Entre as atividades de serviço técnico, encontravam-se: assistência técnica, prestação de serviços, entre outros.

Ainda foi possível extrair dos resultados que 2,9% das pessoas realizaram outras 42 atividades diferentes que não são pré-definidas pelo programa Lattes, como, por exemplo, avaliação de projetos de pesquisa e desenvolvimento, revisão de artigos científicos, participação em exame de qualificação, entre outras.

Além disso, analisando as produções bibliográficas, observou-se que foram publicados 573 artigos, de 1968 até o princípio de 2004, sendo que, a maior parte deles foi publicada em 2001. Dentre esses artigos, 6,4% foram publicados no exterior e a maioria possui de 3 a 5 autores, com alguns possuindo até 8 autores. A Figura 4 mostra um gráfico com a distribuição dos artigos por área do conhecimento. De acordo com o site oficial do CNPq⁴, as grandes áreas do conhecimento são: Ciências Agrárias, Ciências Biológicas, Ciências da Saúde, Ciências Exatas, Ciências Humanas, Ciências Sociais Aplicadas, Engenharias e Linguística - Letras e Artes. Todos os cursos de graduação, Pós-graduação e especialização existentes hoje são classificados entre essas áreas. Pode-se citar algumas dessas classificações, como:

- Ciências Agrárias: Agronomia, Ciência do Solo, Zootecnia, Medicina Veterinária, Ciência de Alimentos, Engenharia de Alimentos, Engenharia Florestal, Engenharia Agrícola, entre outros.
- Ciências Biológicas: Biologia Geral, Genética, Botânica, Zoologia, Ecologia, Bioquímica, Biofísica, Farmacologia, entre outros.
- Ciências da Saúde: Medicina, Odontologia, Enfermagem, Nutrição, Fisioterapia, Educação Física, entre outros.
- Ciências Exatas: Matemática, Física, Ciência da Computação, Sistemas de Informação, Astronomia, Química, GeoCiências, entre outros.

⁴Site <http://www.cnpq.br>, visitado em 26 de junho de 2004.

- Ciências Humanas: Filosofia, Sociologia, Arqueologia, História, Geografia, Psicologia, Educação, Teologia, entre outros.
- Ciências Sociais Aplicadas: Direito, Administração, Economia, Ciências Contábeis, Ciência da Informação, Arquitetura e Urbanismo, Arquivologia, Comunicação, turismo, entre outros.
- Engenharias: Engenharia Civil, Telecomunicações, Engenharia Mecânica, Engenharia Econômica, Engenharia Aeroespacial, Engenharia Médica, entre outros.

Dos 573 artigos publicados, 77% pertencem às Ciências Agrárias, 13% às Ciências Biológicas, 2,3% Ciências da Saúde, 5,4% Ciências Exatas, 0,3% Ciências Humanas, 1,7% Ciências Sociais Aplicadas e 0,3% Engenharias.

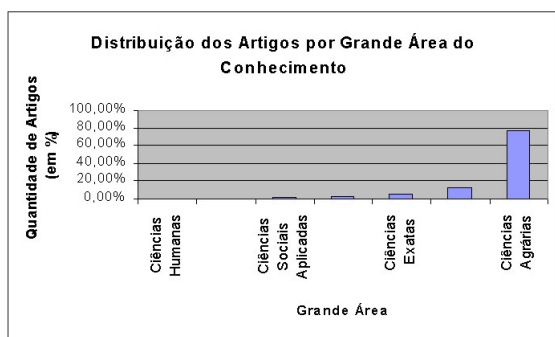


Figura 4: Número de artigos publicados por grande área do conhecimento.

Estes foram alguns dos resultados obtidos com a aplicação do processo de Descoberta de Conhecimento em Banco de Dados. Outros resultados serão obtidos em trabalhos futuros.

8 Considerações Finais

Diante dos resultados apresentados, pode-se perceber claramente que este trabalho conseguiu obter uma visão mais abrangente dos dados institucionais, além de ter disponibilizado uma grande quantidade de informações sobre a Pesquisa Científica da Ufla. É possível iniciar, então, uma

melhoria na Gestão do Conhecimento desta instituição, fazendo uso dessas informações, pois é exatamente essa a base da Gestão do Conhecimento: dados integrados, gerando informações analíticas e abrangentes.

Não se pode negar que as ações empresariais destinadas à busca competitiva no mercado estão baseadas principalmente na Gestão do Conhecimento interno e externo à organização. Não se pode negar, também, que as Universidades, assim como qualquer organização, devem fazer uso da Gestão do Conhecimento para auxiliar a tomada de decisão na busca de uma qualidade de ensino cada vez melhor.

Apesar de possuir algumas limitações, como a desatualização dos currículos e uma certa falta de padronização nos cadastros, este trabalho apresenta-se como uma iniciativa única no sentido de dar uma visão integrada das produções científicas, tecnológicas e bibliográficas de professores e pessoas ligadas à Universidade Federal de Lavras, minerando dados e informações e criando um conhecimento que poderá servir de base para o processo de Gestão da Pesquisa Científica nesta instituição.

Apesar de ter sido aplicado a uma área específica, à Pesquisa Científica na Ufla, o trabalho demonstrou como é possível também utilizar Tecnologias da Informação para auxiliar a Gestão de Instituições de Ensino Superior. Diversos padrões e associações foram identificados através da aplicação da Descoberta de Conhecimento em Banco de Dados, porém, há muitas outras descobertas a serem feitas com o sistema desenvolvido.

Este projeto foi apenas um passo para o desenvolvimento de um grande trabalho de mudança na Gestão do Conhecimento nas atividades gerenciais da Ufla e, quem sabe, futuramente, de outras Universidades. O sistema desenvolvido poderá ser incrementado e utilizado em trabalhos futuros. Entre as atividades que poderão ser realizadas, estão: (1) atualização da Base de Dados a partir da nova versão do Lattes Extrator, (2) entrevistas com pessoas-chave para estabelecer novos critérios de exploração dos dados, (3) desenvolvimento da fase 2 da técnica de *Data Mining* e (4) descoberta de novas informações e novo conhecimento.

Referências

- [Adriaans & Zantinge (1996)] ADRIAANS, P.; ZANTINGE, D. *Data Mining*. 1 ed. Harlow: Addison-Wesley, 1996.
- [Carvalho (2000)] CARVALHO, R. B.. Aplicações de Softwares de Gestão do Conhecimento: Tipologia e Usos. Tese de M.Sc., UFMG, Belo Horizonte, MG, 2000.
- [Decker (2004)] DECKER, K.; FOCARDI, S. Technological Overview: A Report on Data Mining. Disponível em: <ftp://ftp.cscs.ch/pub/CSCS/techreports/1995/CSCS-TR-95-02.ps.gz>, Consultado em 17.mar.2004.
- [Deitel (2003)] DEITEL, H. M., DEITEL, P.J., NIETO, T. R. e SADHU, P. XML Como Programar. Porto Alegre: Bookman, 2003.
- [Elmasri (2002)] ELMASRI, Ramez e NAVATHE, Shamkant B.. *Sistemas de Banco de Dados : Fundamentos e Aplicações*. 3ª Edição. Rio de Janeiro: LTC, 2002.
- [Grossman (2004)] GROSSMAN, R., *Data Mining: challenges and opportunities for Data Mining during the next Decade*. Disponível em: Magnify site: <http://www.magnify.com>, Consultado em 20.mar.2004.
- [King (2004)] KING, D., CS 4803B - Numerical Machine Learning. Disponível em: Georgia Tech College of Computing Site: <http://www.cc.gatech.edu/kingd/datamine/datamine.html> Consultado em 22.mar.2004.
- [Laudon & Laudon (1999)] LAUDON, KENNETH C. e JANE P. *Gerenciamento de Sistema de Informação*, LTC - 1999 - Terceira Edição.
- [Moxton (2004)] MOXTON, B., Defining Data Mining. Disponível em: DBMS Data Warehouse Supplement Site: <http://www.dbms.mfi.com/9608d53.html>. Consultado em 20.mar.2004.
- [Oracle (2004)] ORACLE 9i. *Data Mining: An Oracle White Paper*. Documento disponível no site oficial da Oracle, disponível em <http://www.oracle.com>, consultado em 29.abr.2004.
- [Parrini (2002)] PARRINI, Esther. *Gestão do Conhecimento no Suporte à Decisão OLAP*. Tese de Mestrado. NCE/IM/UFRJ, 2002.
- [Pirola (2002)] PIROLA, Vinícius S.. *A proposição de uma ferramenta de apoio ao Mapeamento do conhecimento em uma organização*. Tese M.Sc., COPPE/UFRJ, Rio de Janeiro 2002.
- [Tachizawa & Andrade (2002)] TACHIZAWA, T., ANDRADE, R. O. B. de. *Gestão de Instituições de Ensino*. 3ª Edição - Rio de Janeiro: Editora FGV, 2002.
- [Tarapanoff (2001)] TARAPANOFF, Kira (organizadora). *Inteligência Organizacional e Competitiva*. Brasília: Editora Universidade de Brasília, 2001.
- [W3C (2004)] W3C - World Wide Web Consortium. *Transformações XSL (XSLT)*. Recomendação de 16 de novembro de 1999. Versão 1.0. Tradução disponível em <http://www.amtechs.com/w3c/rec-xslt-19991116.html>, consultado em 04.mai.2004.