

CRISTHIANE XAVIER FIGUEIREDO

**DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS:
Um Estudo de Caso da Pesquisa Científica na Universidade
Federal de Lavras**

Monografia de Graduação apresentada ao Departamento de Ciência da Computação da Universidade Federal de Lavras como parte das exigências do Curso de Ciência da Computação para obtenção do título de Bacharel

Orientadora
Prof. Olinda Nogueira Paes Cardoso

Lavras
Minas Gerais - Brasil
2004

CRISTHIANE XAVIER FIGUEIREDO

**DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS:
Um Estudo de Caso da Pesquisa Científica na Universidade
Federal de Lavras**

Monografia de Graduação apresentada ao Departamento de Ciência da Computação da Universidade Federal de Lavras como parte das exigências do Curso de Ciência da Computação para obtenção do título de Bacharel

Aprovada em 24 de junho de 2004

André Luiz Zambalde

Prof. Olinda Nogueira Paes Cardoso
(Orientadora)

Lavras
Minas Gerais - Brasil

Agradecimentos

Agradeço, primeiramente, a Deus por mais esta existência de vida, por esta oportunidade de aprendizado e formação intelectual e, principalmente, pela minha saúde, que teve de suportar muitos momentos tristes nessa batalha, mas que também vibrou com várias alegrias e descobertas.

Agradeço a toda minha família, pelo apoio e incentivo a continuar buscando meus ideais; aos meus colegas de sala, pela companhia, amizade e compreensão durante esses quatro anos; aos meus velhos amigos, de quem a distância me priva um pouco; à professora Olinda, orientadora deste projeto, por toda ajuda, toda dedicação e todo ensinamento que ela me passou, não só no período de realização deste trabalho, mas desde que a conheci na faculdade; ao professor Rêmulo, pelas oportunidades de estágio e aprendizado; e, em especial, ao meu namorado, Daniel, por todo carinho, toda força e todo amor;

A todos vocês: Muito Obrigada!

Dedico este trabalho às pessoas mais próximas a mim, que fizeram parte da minha vida estes últimos anos, participando dos bons e maus momentos, perdoando meus defeitos e me proporcionando muitos momentos de alegria: minha mãe, Daniel, Elaine, Luci e Matheus.

Resumo

Descoberta de Conhecimento em Banco de Dados: Um estudo de caso da Pesquisa Científica na Universidade Federal de Lavras

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

Abstract

Knowledge Discovery in Database: A Case Study of Scientific Research in The Federal University of Lavras

This work shows how it is possible to optimize the managerial activity of Universities, it uses the Knowledge Discovery in Databases (KDD) for help in the knowledge Management. Through the application of this technique at Scientific research of the Federal University of Lavras, the present work brought important informations and an integrated vision of all Scientific production, technological and bibliographical of this institution, it is making possible an improvement in knowledge Management in this institution.

Keywords: *data mining, knowledge management, university management*

Sumário

1	Introdução	1
1.1	Visão Geral	1
1.2	Objetivo	3
2	Referencial Teórico	5
2.1	Gestão do Conhecimento	5
2.1.1	Criando Conhecimento	9
2.2	Descoberta de Conhecimento em Banco de Dados	11
2.2.1	Seleção ou Filtragem dos Dados	13
2.2.2	Pré-processamento e Análise dos Dados	14
2.2.3	Transformação dos Dados e Desenvolvimento do Modelo	15
2.2.4	Geração e Interpretação de Resultados	15
2.2.5	Resumo do Processo de DCBD	16
2.3	Data Mining	16
2.3.1	Objetivos do <i>Data Mining</i>	18
2.3.2	Tipos de conhecimento descobertos pelo <i>Data Mining</i> . .	19
2.4	Gestão de Universidades	20
2.5	A Pesquisa Científica na Ufla	23
2.6	Plataforma Lattes	29
2.6.1	Lattes Extrator	29
3	Metodologia	31
3.1	Tipos de Pesquisa	31
3.2	Procedimento Metodológico	32
3.3	Desenvolvimento	33
4	Resultados e Discussões	45

5	Considerações Finais	53
A	Tabelas do BD Lattes	57
B	Código Fonte XSL	59

Lista de Figuras

2.1	Modelo genérico de Gestão do Conhecimento	8
2.2	Os quatro processos de conversão do conhecimento	10
2.3	Processo de Descoberta de Conhecimento em Banco de Dados . .	13
2.4	Exemplo de Regra de Associação.	19
3.1	Interface Web do Lattes Extrator.	32
3.2	Execução das etapas do processo de DCBC neste trabalho.	34
3.3	Parte do código XML utilizado na conversão.	37
3.4	Parte do código XSL utilizado na conversão.	39
3.5	Parte do script SQL gerado a partir da conversão.	40
3.6	Exemplo do código de algumas consultas SQL feitas ao Banco de Dados deste trabalho.	42
4.1	Número de atuações profissionais por pessoa.	46
4.2	Número de atividades de ensino por pessoa.	46
4.3	Número de atividades de direção por pessoa.	47
4.4	Número de atividades de pesquisa por pessoa.	48
4.5	Número de atividades de extensão por pessoa.	48
4.6	Número de serviços técnicos por pessoa.	49
4.7	Número de treinamentos ministrados por pessoa.	49
4.8	Número de artigos publicados por ano.	50
4.9	Número de autores por artigo.	51
4.10	Número de artigos publicados por grande área do conhecimento. .	52

Capítulo 1

Introdução

1.1 Visão Geral

Nos dias atuais, pode-se observar que as organizações vêm sofrendo uma grande mudança na forma de gerir recursos materiais, pessoas e, principalmente informações. Isto se deve ao surgimento de novas tecnologias para gerenciamento de informação, ao novo ambiente empresarial - dinâmico, aberto e competitivo -, às novas formas de organização das empresas - abertas com atuação em rede fundamentada na informação - e à nova ordem geopolítica - realidade mundial aberta, volátil e multipolar.

Essas mudanças convergiram, então, para a quebra de um paradigma histórico e permitiu a entrada de uma nova era chamada Sociedade da Informação e do Conhecimento, em que a informação constitui a principal matéria-prima, o conhecimento é utilizado na agregação de valor de produtos e serviços, a tecnologia constitui um elemento vital para as mudanças e a rapidez, a efetividade e a qualidade constituem fatores decisivos de competitividade.

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, tornando possíveis ações inteligentes, inovações e capacidade de continuamente criar produtos e serviços excelentes em termos de complexidade, flexibilidade e criatividade.

Neste cenário, 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.

O processo de Gestão do Conhecimento abrange toda forma de gerar, armazenar, distribuir e utilizar o conhecimento, tornando necessária a utilização de Tecnologias da Informação para facilitar este processo. Assim, o desenvolvimento de Tecnologias da Informação, principalmente no que se refere à geração e ao armazenamento de dados, possibilita registrar qualquer transação realizada em grandes bases de dados sempre crescentes. Mas a coleta e o armazenamento de dados por si só não contribuem para alavancar a estratégia da empresa. É necessário que se façam análises sobre essa grande quantidade de dados, estabelecendo indicadores para descobrir relações de causa e efeito, pois processar informação correta é um dos requisitos mais essenciais para uma boa tomada de decisão.

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 de 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. Para que o conhecimento seja descoberto, técnicas de *Data Mining* (mineração de dados) devem ser aplicadas.

Data Mining é uma técnica que faz parte de uma das etapas da Descoberta de Conhecimento em Banco de Dados. Essa técnica é capaz de revelar o conhecimento que está implícito em grandes quantidades de informações armazenadas nos bancos de dados de uma organização. Esta técnica pode fazer uma análise antecipada dos eventos, possibilitando prever tendências e comportamentos futuros, permitindo aos gestores, tomar decisões baseadas em fatos e não em suposições.

Assim como as empresas, as Universidades precisam inovar, otimizando a atividade gerencial, principalmente, para acompanhar as exigências do mercado, que está em constante modificação e que necessita cada vez mais de profissionais capacitados a lidar com os novos paradigmas e com as tecnologias emergentes. Nunca se falou tanto em avaliar a qualidade dos serviços prestados pelas Instituições de Ensino Superior (IESs) e nunca se questionou tanto a qualidade e os valores cobrados por esses serviços. Tem-se acentuado a necessidade de se repensar as IESs, preparando-as para as transformações que estão ocorrendo no ambiente em que operam.

Como qualquer organização, as Universidades têm por objetivo principal, prestar serviços de ensino, de pesquisa e de extensão de alta qualidade e, concomitantemente, assegurar um bom ambiente de trabalho para os seus funcionários. As IESs não podem mais ficar estacionárias no que diz respeito às características e

fatores que lhe conferem uma vantagem competitiva. Cada qual tem que inovar incessantemente para poder competir e sobreviver. Por que, então, não acompanhar as empresas na implantação e utilização de novas tecnologias que auxiliem o gestor das Universidades a fazer frente aos novos tempos?

Levando em consideração essas questões e alguns problemas enfrentados pelas Universidades como o gerenciamento dos dados, a visão extremamente segmentada, setorizada que a maioria tem delas mesmas e as diversas limitações encontradas na gestão dos sistemas de informações, o presente trabalho utilizou a técnica de *Data Mining* para extrair informações importantes que servem de apoio à solução de problemas, à tomada de decisão e à Gestão do Conhecimento na Universidade Federal de Lavras (Ufla).

Como a Gestão do Conhecimento constitui um processo que parte da integração das informações de todas as áreas da organização, foi realizado um estudo aprofundado do uso das informações da Ufla, em especial, as informações referentes ao corpo docente desta instituição. Tendo em vista que é possível extrair um grande número de informações úteis a partir da análise das produções científica, tecnológica e bibliográfica dos professores, foi criada uma base de dados a partir de arquivos extraídos da Plataforma Lattes e, posteriormente, aplicada a técnica de Descoberta de Conhecimento em Banco de Dados, identificando padrões, associações e tendências, e gerando base para a Gestão do Conhecimento nesta Universidade.

1.2 Objetivo

Este trabalho tem como objetivo geral desenvolver e aplicar o processo de Descoberta de Conhecimento em Banco de Dados para extrair conhecimento referente à Produção Científica dos professores da Ufla. Como objetivos específicos, tem-se:

1. Selecionar e tratar os dados disponíveis na Plataforma Lattes referentes à pesquisa científica na Ufla;
2. Implementar um sistema para importar os dados citados acima para um Banco de Dados;
3. Definir os principais problemas gerenciais relacionados à pesquisa científica na Ufla;
4. Desenvolver uma ferramenta automática de descoberta de conhecimento, utilizando a técnica de *Data Mining*;

5. Avaliar a ferramenta e as informações geradas.

Esta monografia está organizada como se segue: o Capítulo 2 apresenta várias definições que serão necessárias para o entendimento da realização do projeto. Entre essas definições, temos: Gestão do Conhecimento, com seus objetivos e alguns procedimentos adotados; Descoberta de Conhecimento em Bancos de Dados; *Data Mining*, sua definição, seu funcionamento básico e suas principais etapas; Gestão de Universidades, seus principais aspectos; a Linguagem de marcação XML ¹; e, por fim, serão abordados alguns pontos críticos da política de gestão da Pesquisa Científica na Universidade Federal de Lavras.

O Capítulo 3 apresenta a metodologia adotada para a realização deste projeto, bem como que atividades e ferramentas foram realizadas para viabilização do mesmo e como ele foi realizado. No Capítulo 4, temos os resultados e algumas discussões sobre esses resultados. E no Capítulo 5, as conclusões tiradas de todo o processo de desenvolvimento deste projeto.

¹A XML (*Extensible Markup Language*) é uma linguagem de marcação utilizada em documentos da Internet e que apresenta as informações inseridas nos marcadores da linguagem

Capítulo 2

Referencial Teórico

2.1 Gestão do Conhecimento

A velocidade de mudanças em nossa sociedade e o aumento da competição dos mercados globais tem contribuído para um processo de questionamento de quais seriam os pilares fundamentais para o sucesso das organizações. Além disso, o ciclo de desenvolvimento de produtos e serviços nas empresas tem sido drasticamente reduzido e as organizações buscam cada vez mais qualidade, inovação e velocidade para permanecerem no mercado. Para sobreviver, as organizações precisam aprender a diferenciar seus produtos e serviços através do conhecimento [Carvalho (2000)].

Pirolla (2002), citando Davenport & Prusak (1998), afirma que neste novo contexto de negócios, as atividades baseadas no conhecimento, como o desenvolvimento de novos processos e produtos, estão se tornando primordiais para as empresas. As corporações estão se diferenciando umas das outras pelo tanto pelo que produzem, mas pelo que sabem.

Nesta mesma linha de raciocínio, Pirolla (2002) cita Barroso & Gomes (1999), afirmando que, em um mercado cada vez mais competitivo o sucesso nos negócios depende basicamente da qualidade do conhecimento que cada organização aplica nos seus processos corporativos/empresariais. Neste contexto, o desafio de utilizar o conhecimento residente na empresa com o objetivo de criar vantagens competitivas torna-se crucial.

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). Antes de chegar a uma definição do que seja gerenciar o Conhecimento, é necessário conceituar “conhecimento”. Vários autores fazem uma distinção ascendente entre dado, informação e conhecimento.

- Dados: são fatos, imagens ou sons que podem ou não ser úteis ou pertinentes para uma atividade particular. São abstrações formais quantificadas, que podem ser armazenadas e processadas por computador.
- Informações: são dados contextualizados com forma e conteúdo apropriados para um uso particular. São abstrações informais (não podem ser formalizadas segundo uma teoria matemática ou lógica) que representam, por meio de palavras, sons ou imagens, algum significado para alguém.
- Conhecimento: é uma combinação de instintos, idéias, informações, regras e procedimentos que guiam ações e decisões; têm embutido em si valores como sabedoria e insights. É a inteligência obtida pela experiência. Como exemplo pode-se citar a experiência que um funcionário possui por ter trabalhado em determinadas atividades numa organização por muito tempo.

O conhecimento organizacional pode ser classificado em dois tipos, segundo Tarapanoff (2001). Um é o conhecimento explícito, que pode ser articulado na linguagem formal, sobretudo em afirmações gramaticais, expressões matemáticas, especificações, manuais e assim por diante. Esse tipo de conhecimento pode ser então transmitido, formal e facilmente, entre os indivíduos.

O segundo tipo, o conhecimento tácito, é difícil de ser articulado na linguagem formal. É o conhecimento pessoal, incorporado à experiência individual e envolve fatores intangíveis como, por exemplo, crenças pessoais, perspectivas e sistemas de valor. O conhecimento tácito foi deixado de lado como componente crítico do comportamento humano coletivo. A dimensão cognitiva do conhecimento tácito reflete nossa imagem da realidade (o que é) e nossa visão do futuro (o que deveria ser). Apesar de não poderem ser articulados muito facilmente, esses

modelos implícitos moldam a forma com que percebemos o mundo à nossa volta [Tarapanoff (2001)].

Podemos considerar o conhecimento explícito e o conhecimento tácito como unidades estruturais básicas que se complementam. Mais importante, a interação entre essas duas formas de conhecimento é a principal dinâmica da criação do conhecimento em uma organização. A criação do conhecimento organizacional é um processo em espiral em que a interação ocorre repetidamente [Tarapanoff (2001)].

Na medida em que o conhecimento - tanto o tácito quanto o explícito - se torna um ativo central, produtivo e estratégico, o sucesso da organização depende cada vez mais da sua habilidade em coletar, produzir, manter e distribuir conhecimento.

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)].

Parrini (2002), citando Malhotra (1998), afirma que a Gestão do Conhecimento serve como instrumento para adaptação, sobrevivência e competência organizacional em face de crescentes mudanças ambientais descontínuas. Em sua essência, a Gestão do Conhecimento engloba processos organizacionais que buscam a combinação sinérgica de dados e capacidade de processamento das tecnologias da informação, além da capacidade criativa e inovativa dos seres humanos.

Em outras palavras, a Gestão do Conhecimento é um conjunto de processos, apoiados por ferramentas de Tecnologia da Informação, voltados a capturar, organizar, armazenar, proteger e - sobretudo - compartilhar o conhecimento das pessoas, sob suas duas formas: conhecimento explícito (dados e informações) e conhecimento tácito (habilidades e experiências).

Considerando ainda a definição de Beckman (1999), citada por Tarapanoff (2001), temos que Gestão do Conhecimento é a formalização das experiências, conhecimentos e expertise, de forma que se tornem acessíveis para a organização,

e esta possa criar novas competências, alcançar desempenho superior, estimular a inovação e criar valor para seus clientes.

Existem atualmente vários modelos de Gestão do Conhecimento que se diferenciam, basicamente, em algumas especificidades e na aplicabilidade. Porém, existem idéias básicas que permeiam todos eles, permitindo a definição de um modelo genérico, como mostrado na Figura 2.1.

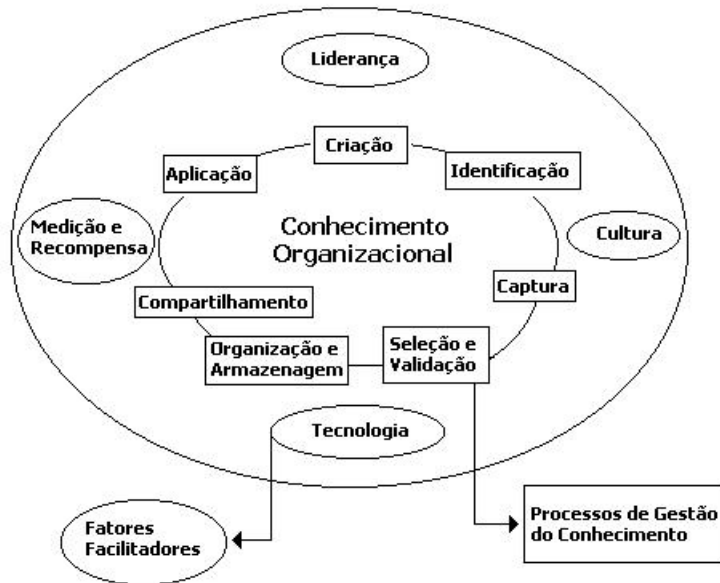


Figura 2.1: Modelo genérico de Gestão do Conhecimento

Sete processos foram considerados por Tarapanoff (2001), na composição do modelo genérico de gestão do conhecimento:

1. Identificação - reconhecer que competências são críticas para o melhor desempenho da organização (competências essenciais);
2. Captura - adquirir conhecimentos, habilidades e experiências necessárias para criar e manter as competências essenciais e áreas de conhecimento selecionadas e mapeadas;

3. Seleção e Validação - filtrar o conhecimento, avaliar sua qualidade e sintetizá-lo para fins de aplicação futura (fortemente associada ao processo de captura);
4. Organização e armazenagem - refletir sobre algumas questões consideradas básicas, como que conhecimento a organização quer ou de ver guardar, de que conhecimento a organização necessita, que conhecimento deve ser ignorado ou descartado, e qual a melhor forma de recuperar o conhecimento.
5. Compartilhamento (acesso e distribuição) - organizar e formalizar o conhecimento para que seja armazenado eletronicamente, tornando-o disponível em qualquer parte, a qualquer tempo e em qualquer formato.
6. Aplicação - aplicar em situações reais da organização os conhecimentos, as experiências e informações disponíveis, de modo a produzir benefícios concretos, como melhoria no desempenho, lançamento de novos produtos e conquista de novos mercados.
7. Criação do Conhecimento - envolve as seguintes dimensões: aprendizagem, externalização do conhecimento, lições aprendidas, pensamento criativo, pesquisa, experimentação, descoberta e inovação.

2.1.1 Criando Conhecimento

A criação de conhecimento organizacional pode ser definida como a capacidade que uma instituição tem de criar conhecimento, disseminá-lo na organização e incorporá-lo a produtos, serviços e sistemas. Criar novos conhecimentos também não é apenas uma questão de aprender com os outros ou adquirir conhecimentos externos. O conhecimento deve ser construído por si mesmo, muitas vezes exigindo uma interação intensiva e laboriosa entre diversos membros da organização.

Assim, diz respeito também tanto aos ideais como às idéias. Ele também pode ser definido na hora (“aqui e agora”) com base na experiência direta e por meio da tentativa e erro, o que exige intensa e trabalhosa interação entre os membros da equipe [Tarapanoff (2001)].

As formas de interação entre o conhecimento tácito e o conhecimento explícito, e entre o indivíduo e a organização, realizarão quatro processos principais da conversão do conhecimento, que, juntos, constituem a criação do conhecimento, segundo a afirmação de Tarapanoff (2001), ao citar Nonaka & Takeuchi (1997). A Figura 2.2 apresenta uma ilustração desses quatro processos, que são os seguintes:

- Do Tácito para o explícito (externalização), que é um processo de articulação do conhecimento tácito em conceitos explícitos, ou seja, de criação do conhecimento perfeito, à medida que o conhecimento tácito se torna explícito, expresso na forma de analogias, conceitos, hipóteses ou modelos;
- Do explícito para o explícito (combinação), cujo modo de conversão do conhecimento envolve a combinação de conjuntos diferentes de conhecimento explícito;
- Do explícito para o tácito (internalização), que é o processo de incorporação do conhecimento explícito no conhecimento tácito;
- Do tácito para o tácito (socialização), que é um processo de compartilhamento de experiências e, a partir daí, de criação do conhecimento tácito, como modelos mentais ou habilidades técnicas compartilhadas.



Figura 2.2: Os quatro processos de conversão do conhecimento
 Fonte: Tarapanoff (2001)

Para criação de conhecimento explícito, diversas técnicas de Descoberta de Conhecimento podem ser utilizadas pelas organizações. Um dos maiores problemas enfrentados atualmente é o grande volume das bases de dados das empresas.

A Descoberta de Conhecimento em Banco de Dados pode ser utilizada como solução para este problema.

2.2 Descoberta de Conhecimento em Banco de Dados

A crescente necessidade de informações disponíveis vem crescendo assustadoramente nos últimos anos e vários fatores contribuíram para este incrível aumento. O baixo custo na armazenagem de pode ser vista como a principal causa do surgimento destas enormes bases de dados. Um outro fator é a disponibilidade de computadores de alto desempenho a um custo razoável. Como consequência, bancos de dados passam a conter verdadeiros tesouros de informação e, devido ao seu volume, ultrapassam a habilidade técnica e a capacidade humana na sua captação e interpretação.

É preciso transformar esses dados armazenados em informação para que esta seja um instrumento estratégico de apoio às tomadas de decisão, podendo ajudar a melhorar procedimentos, detectar tendências e características disfarçadas, e até prevenir ou reagir a um evento que ainda está por vir.

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 algoritmos ¹ para a extração de padrões em um conjunto de dados, mas sem os passos adicionais em um processo de descobrimento do conhecimento. Uma definição formal é que DCBD é o processo não trivial de identificação de padrões em um conjunto de dados com as seguintes características:

- Validade: A descoberta de padrões deve ser válida em novos dados com algum grau de certeza ou probabilidade.
- Novidade: Os padrões são novos (pelo menos para o sistema em estudo), ou seja, ainda não foram detectados por nenhuma outra abordagem.
- Utilidade Potencial: Os padrões devem poder ser utilizados para a tomada de decisões úteis, medidas por alguma função.
- Assimiláveis: Um dos objetivos do DCBD é tornar os padrões assimiláveis ao conhecimento humano.

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.

O processo de DCBD é interativo, envolvendo inúmeras tarefas com muitas decisões tomadas pelo usuário. O analista envolvido em um processo de Descoberta de Conhecimento, em resposta a um determinado objetivo, extrai de um banco de dados, através de uma consulta, um conjunto de dados para sua análise. Após a geração desse conjunto de dados, ferramentas de análises e visualização são utilizadas. Essas análises levam ao analista algumas informações preliminares sobre as questões relacionadas com o objetivo. Essas informações são apresentadas e difundidas na organização.

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

¹Algoritmo é uma seqüência de passos destinados a resolver um problema ou atingir algum objetivo específico

Interpretação de Resultados. Em seguida, cada uma dessas etapas será observada com maior detalhe. A Figura 2.3 representa estas etapas do processo de DCBD.

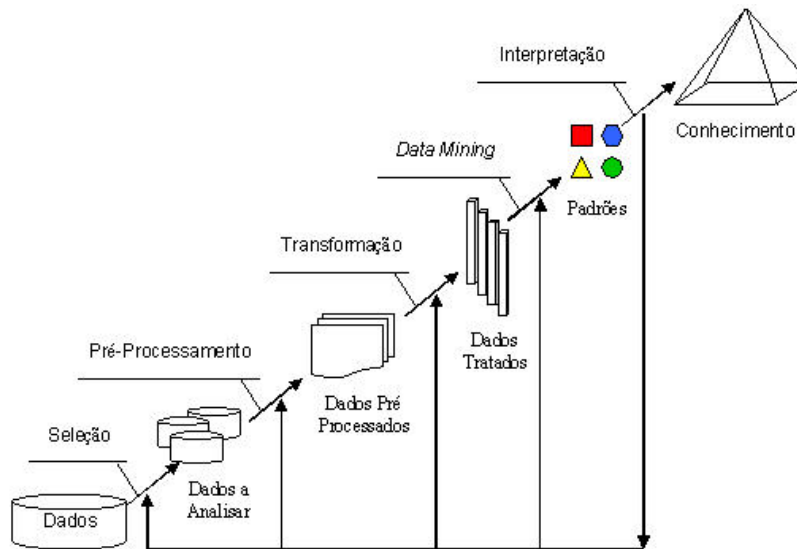


Figura 2.3: Processo de Descoberta de Conhecimento em Banco de Dados
 Fonte: Santos, M. F.. Descoberta de Conhecimento em Bases de Dados.

2.2.1 Seleção ou Filtragem dos Dados

Como os dados são coletados de diversas fontes, muitos erros, arquivos repetidos ou com dados nulos devem ser excluídos do conjunto de dados a ser estudado. Um processo de DCBD não pode ter sucesso sem uma etapa inicial de filtragem ou limpeza dos dados.

O método mais comum, utilizado para verificar a consistência de um conjunto de dados, é selecionar o mesmo dado de fontes múltiplas e comparar seus resultados. No entanto, é uma tarefa que exige um conhecimento tácito muito grande do analista, pois ele deve discernir entre um dado realmente incorreto e uma exceção (*outlier*) à regularidade do conjunto de dados que deve ser incluída na amostra.

Dois problemas principais de incongruência no banco de dados podem ocorrer: dados duplicados e inconsistência no domínio. Dados duplicados são aqueles que trazem a mesma informação, sendo diferenciados por algum erro na entrada do arquivo (por exemplo, erro na digitação do nome de uma pessoa). Existem

ferramentas de busca por arquivos duplicados através da semelhança em seu registro. Além disso, o problema de filtragem dos dados pode ser dinâmico, ou seja, se o processo de DCBD for utilizar dados coletados continuamente e existir alguma falha na coleta, um processo sistemático de filtragem desses dados deve ser implementado. Não apenas uma vez, mas continuamente.

2.2.2 Pré-processamento e Análise dos Dados

O analista em geral possui uma hipótese sobre o conjunto de dados e algum tipo de ferramenta de análise é utilizado para a construção de um modelo. Em geral, a idéia é entender porque certos grupos de entidades comportam-se de certo modo.

Os processos principais na Análise de Dados são:

- Especificação do modelo: um modelo específico é escrito de uma maneira formal;
- Ajuste do modelo: alguns parâmetros específicos do modelo são determinados, quando necessário, de acordo com o conjunto de dados;
- Avaliação do modelo: o modelo é avaliado com o conjunto de dados, através de um conjunto de teste, onde os valores de entrada e saída são conhecidos previamente;
- Refino do modelo: o modelo inicial é iterativamente alterado até que algum parâmetro de erro seja alcançado.

Em ferramentas de análise baseadas em algoritmos, um modelo é especificado através da associação de variáveis de entrada (independente) e variáveis de saída (dependente). Em certos casos, o conjunto de dados é subdividido em um conjunto de dados para treinamento e um conjunto de dados para teste. O conjunto de dados de treinamento é utilizado para ajustar os parâmetros do modelo. O modelo é então avaliado através de sua aplicação no conjunto de teste.

Em ferramentas de análises baseadas em visualização, a hipótese é especificada através da visualização do conjunto de dados e da seleção de elementos nos dados. A própria visualização produzida é o modelo, e seu poder explicativo pode ser observado através da visualização. A visualização é um ingrediente chave para as três principais tarefas no processo de DCBD: desenvolvimento do modelo, análise de dados e geração de resultados.

A representação apropriada dos pontos de dados e suas relações podem dar ao analista informações preliminares que não são possíveis de serem obtidas através

de tabelas de resumos estatísticos. Em alguns casos, a visualização dos dados é a única solução para problemas de confirmação de hipóteses.

Em qualquer problema real de descoberta de conhecimento, o analista necessita utilizar ferramentas baseadas em algoritmos e visualização, iterativamente. Os resultados do uso de uma ferramenta são utilizados para refinar as entradas para outra ferramenta.

2.2.3 Transformação dos Dados e Desenvolvimento do Modelo

Raramente um projeto inicia-se com a hipótese já definida. Uma das operações principais é descobrir subconjuntos da população que se comportem de forma semelhante no foco da análise. Em muitos casos, a população inteira pode ser muito diversa para compreensão, mas detalhes dos subconjuntos podem ser trabalhados.

A interação com o conjunto de dados leva a formulação das hipóteses. Nessa fase, os três principais sub-processos são: (1) segmentação dos dados, (2) seleção do modelo de *Data Mining* e (3) seleção de parâmetros. Para a segmentação dos dados, podem ser utilizadas ferramentas de *clustering*. Para a seleção do modelo de *Data Mining* uma grande variedade de modelos de análises podem ser utilizados, como regressão, árvores de decisão, redes neurais e regras de associação. O analista deve escolher o melhor tipo de modelo antes de iniciar a utilização de uma ferramenta específica. As fases de análise e desenvolvimento do modelo são complementares, e o analista pode voltar e alterar cada fase iterativamente. Esse ciclo é crucial ao processo de descoberta.

2.2.4 Geração e Interpretação de Resultados

Num cenário simplista, uma análise resulta em um relatório de algum tipo que pode incluir medidas estatísticas do modelo, dados sobre exceções, etc. Em geral, os resultados devem ser gerados de forma variada e simplificada. Uma descrição textual de uma tendência ou um gráfico que capture as relações no modelo são mais apropriados.

Ações também são indicadas, ou seja, gerar e detalhar procedimentos que o usuário deva tomar dependendo de certas características. O resultado de um processo de DCBD deve ser visto como uma especificação de uma aplicação a ser construída que responda às questões chave sobre o consumidor. Isso usualmente vem na forma de um modelo que será utilizado como uma especificação principal para uma aplicação.

2.2.5 Resumo do Processo de DCBD

O processo de DCBD se inicia com a identificação dos objetivos do estudo, ou seja, quais informações devem ser obtidas do banco de dados. Em seguida, iniciam-se as quatro etapas descritas anteriormente, com a utilização de diversas ferramentas auxiliares. Por fim, após a geração dos resultados, temos como produtos um relatório, com as principais informações definidas nos objetivos, a necessidade de implementar ações a partir das informações obtidas, a aplicação do modelo desenvolvido em outras áreas e o monitoramento desse processo de implementação.

Através dos pontos observados anteriormente, podemos traçar a relação entre os conceitos apresentados. O *Data Mining* está inserido numa metodologia que procura uma descrição lógica ou matemática, eventualmente de natureza complexa, de padrões e associações em um conjunto de dados.

2.3 Data Mining

A temática de *Data Mining* é bastante recente, iniciando-se na década de 90. Embora muita informação já exista sobre o tema, não existe uma padronização e classificação universalmente aceita sobre o assunto, de maneira a facilitar os interessados da área na condução de seus projetos de pesquisa. Uma das justificativas é justamente essa dimensão de novidade do tema e sua relevância na solução para análise de grandes volumes de dados.

Além disso, o material existente sobre *Data Mining* possui abordagens heterogêneas, dependendo da origem ou do público alvo ao qual se destina. O tema é estudado e abordado por profissionais da área de estatística, informática, matemática, administração, entre outras. Cada tipo de área possui abordagens específicas, adequadas para as suas necessidades.

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 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.

A aplicação típica de *Data Mining* começa com um grande conjunto de dados e poucas definições. A maioria dos algoritmos trata os dados iniciais como uma “caixa-preta”, com nenhuma informação disponível sobre o que os dados descrevem, quais relações existem entre os dados e se contém erros. Ao examinar os dados, um algoritmo pode explorar milhares de prováveis regras, utilizando diversas técnicas para escolher entre elas.

Decker & Focardi (2004) definem *Data Mining* como uma metodologia que procura uma descrição lógica ou matemática, eventualmente de natureza complexa, de padrões e regularidades em um conjunto de dados.

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.

De acordo com Bruce Moxton (2004), *Data Mining* é um conjunto de técnicas utilizadas para explorar exhaustivamente e trazer à superfície relações complexas em um conjunto grande de dados. Uma diferença significativa entre as técnicas de *Data Mining* e outras ferramentas analíticas é a abordagem utilizada para explorar as inter-relações entre os dados. Semelhantemente à abordagem dada por Grossman, também diferencia as técnicas de *Data Mining* com relação às técnicas analíticas entre a abordagem de suposição e a abordagem de descoberta. Segundo ele, as técnicas de *Data Mining* não pressupõem que as relações entre os dados devam ser conhecidas a priori.

Em outras palavras, o *Data Mining* nada mais é do que um algoritmo matemático utilizado para produzir conhecimento 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).

As grandes perguntas que todos os executivos desejam responder de forma rápida são: O que querem nossos clientes? Como anda a concorrência? Que assuntos causam mais impacto na sociedade? Qual o direcionamento de nosso orçamento? Como obter o máximo de informação útil para minha empresa? Como descobrir padrões de dados e novos conhecimentos? Como manter meu cliente? Como utilizar adequadamente e descobrir ligações entre eventos nas minhas bases de dados?

É para encontrar as respostas para essas perguntas que as empresas estão utilizando o *Data Mining*. Essa técnica possibilita prever tendências e comportamentos futuros, permitindo aos gestores tomarem decisões baseadas em fatos e não em suposições. A análise automatizada e antecipada oferecida pelo *Data Mining*, vai muito além da simples análise de eventos passados, que é fornecida pelas ferramentas de retrospectiva típicas de sistemas de apoio à decisão.

Com a utilização da técnica de *Data Mining* novas informações de cunho explícito podem ser geradas. Tais informações podem fazer parte do conjunto de conhecimentos explícitos de uma organização, podendo servir de subsídio para gerar insights e elementos para conhecimento tácito.

2.3.1 Objetivos do *Data Mining*

O objetivo do *Data Mining* é descobrir, de forma automática ou semi-automática, o conhecimento que está “escondido” nas grandes quantidades de informações armazenadas nos bancos de dados da organização, permitindo agilidade na tomada de decisão. Uma empresa que emprega a técnica de *Data Mining* é capaz de:

- Criar parâmetros para entender o comportamento do consumidor;
- Identificar afinidades entre as escolhas de produtos e serviços;
- Prever hábitos de compras;
- Analisar comportamentos habituais para se detectar fraudes.

Em termos gerais, segundo Elmasri & Navathe (2002), a técnica de *Data Mining* compreende os seguintes propósitos:

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.

Previsão - o *Data Mining* pode mostrar como certos atributos dentro dos dados irão comportar-se no futuro.

Identificação - Padrões de dados podem ser utilizados para identificar a existência de um item, um evento ou uma atividade. Envolve uma comparação de parâmetros ou imagens ou sinais em relação ao Banco de Dados.

Classificação - o *Data Mining* pode repartir os dados de modo que diferentes classes ou categorias possam ser identificadas com base em combinações de parâmetros. Às vezes, a classificação baseada no conhecimento sobre o domínio comum é utilizada como uma entrada para decompor o problema de garimpagem e torná-lo mais simples.

Otimização - otimizar o uso de recursos limitados como tempo, espaço, dinheiro ou matéria-prima e maximizar variáveis de resultado como vendas ou lucros sob um determinado conjunto de restrições.

2.3.2 Tipos de conhecimento descobertos pelo *Data Mining*

Segundo Tarapanoff (2001) e Elmasri & Navathe (2002), o conhecimento descoberto durante a fase de *Data Mining* pode ser descrito de cinco maneiras:

1. **Regras associativas** - encontrar itens que determinem a presença de outros em uma mesma transação, ou seja, estabelecer regras que correlacionam a presença de um conjunto de itens com um outro intervalo de valores para um outro conjunto de variáveis. Exemplo ilustrado na Figura 2.4: sempre que se orienta um aluno de doutorado, é publicado algum documento;



Figura 2.4: Exemplo de Regra de Associação.

2. **Hierarquias de classificação** - criar um modelo baseado em dados conhecidos, ou seja, a partir de um conjunto existente de eventos ou transações definir uma hierarquia de classes. Exemplo: os professores de uma instituição podem ser divididos em escalas de acordo com o tempo de serviço.
3. **Padrões Seqüenciais** - encontrar padrões ou comportamento previsível em um período de tempo. Isso significa que um comportamento particular

em um dado momento pode ter como consequência outro comportamento ou seqüência de comportamentos dentro de um mesmo período de tempo. Exemplo: uma pessoa que cursou doutorado no exterior provavelmente publicará no exterior.

4. **Padrões em séries Temporais** - obter todas as ocorrências de subsequências similares em uma base de dados, ou seja, detectar semelhanças dentro de posições de séries temporais. Exemplo: a cada semestre são inseridas novas atividades de ensino.
5. **Categorização e Segmentação** - agrupar registros que contêm características similares, ou seja, particionar (segmentar) uma dada população de eventos ou itens em conjuntos. Exemplo: as pessoas podem ser divididas em categorias de acordo com seu tipo de envolvimento com a Universidade (professores, alunos, técnicos, etc).

O *Data Mining* usa ferramentas de análise estatística, assim como técnicas da área de Inteligência Artificial, como redes neurais, lógica fuzzy, algoritmos genéticos ou técnicas baseadas em regras e outras técnicas inteligentes. A mineração dos dados pode dar-se sobre um banco de dados operacional, ou sobre um *Data Warehouse*, constituindo um Sistema de Suporte à Decisão ³.

A área da mineração de dados e descoberta de conhecimento ainda é uma área relativamente nova, mas as ferramentas de *Data Mining* estão evoluindo constantemente, incorporando, por exemplo, algoritmos extraídos da Inteligência Artificial, da estatística e da otimização.

2.4 Gestão de Universidades

A gestão de uma instituição de ensino típica é um conjunto de decisões assumidas a fim de obter um equilíbrio dinâmico entre missão, objetivos, meios e atividades acadêmicas e administrativas [Tachizawa & Andrade (2002)].

A compreensão da instituição de ensino como um todo e da sua inter-relação com os demais agentes do ramo de atividades - setor educacional - ao qual pertence é essencial para se desenvolver uma proposta de ferramenta de auxílio à gestão do conhecimento, objetivo geral desse projeto. Faz-se necessário analisar finalidades

³Sistemas que dão suporte ao nível gerencial da organização. Possuem poder analítico para ajudar na solução de problemas que não podem ser previstos com antecedência.

e missão, bem como identificar produtos, mercados, fornecedores, concorrentes e órgãos normativos oficiais.

Tal compreensão permitirá estabelecer traços comuns a uma IES e também delinear estratégias genéricas inerentes a uma instituição de ensino típica.

Tachizawa & Andrade (2002) fazem um questionamento acerca da visão que se tem a respeito das IES. Citando Fernandes (1998), considera a Universidade uma empresa prestadora de serviço que oferece produtos. Estes seriam os profissionais formados, capazes de se inserir no âmbito de trabalho e na sociedade em geral.

Por fornecedores entende-se as entidades/agentes que fornecem recursos às IES na forma de bens, capital, materiais, equipamentos e demais recursos, que por sua natureza constituem os insumos necessários às atividades internas das instituições de ensino. Nesse contexto, a figura do professor surge como o principal fornecedor (colaborador ou parceiro) da IES [Tachizawa & Andrade (2002)].

Considerando que o produto final de uma IES é o aluno formado, capacitado e habilitado a exercer a profissão para a qual se preparou, o cliente é a organização empregadora desse profissional colocado no mercado. Mercado, por sua vez, compreende o conjunto de clientes, constituído das organizações que potencialmente irão absorver os profissionais formados e colocados disponíveis pelas instituições de ensino.

À medida que o gestor de IES tem êxito em integrar o cliente e unir os interesses deste aos objetivos preestabelecidos no plano estratégico (projeto pedagógico) da instituição de ensino, refluiriam os resultados que assegurariam o cumprimento da missão, e sobretudo, a sobrevivência (continuidade). São esses resultados que de fato importam à comunidade como um todo e ao gestor da IES em particular [Tachizawa & Andrade (2002)].

Nesta caracterização de uma IES, Tachizawa & Andrade (2002) ainda enfocam alguns elementos de análise, como:

Missão - em que se procura explicitar a finalidade peculiar que diferencia a instituição de ensino de outras do seu tipo;

Produtos e Processos - o que envolve produtos principais, complementares, substitutos e produtos concorrentes a partir da análise dos seguintes fatores:

- Grau de homogeneidade ou heterogeneidade dos produtos gerados pelas IES;
- Qualidade do produto, pesquisas e desenvolvimento;
- Processos produtivos e tecnologia educacional instalada;

- Imagem inerente ao composto de produtos da instituição de ensino;
- Inovação tecnológica decorrente de investimentos em desenvolvimento pedagógico e acadêmico;
- Possibilidade de aquisição de tecnologias educacionais como meio de obtenção de posicionamento competitivo.

Mercado - em que se procura estabelecer a forma de prestação de serviços educacionais, definindo se são feitas diretamente para os clientes ou através de intermediários.

Fornecedores - faz-se o mapeamento dos professores existentes no mercado que, potencialmente, sejam úteis à instituição de ensino.

Concorrentes - procura-se identificar sua origem, características, pontos fortes, e pontos fracos.

Ramo de Atividades - identifica-se qual tipo de setor econômico a que a instituição de ensino sob estudo pertence.

Constata-se, a partir dessas análises e das observações de Tachizawa & Andrade (2002), que é imprescindível agrupar organizações, dentre elas as IES, que, genericamente, têm características similares, para verificar o funcionamento de blocos de organizações e o comportamento das forças competitivas dentro de cada bloco.

Apesar dessa analogia feita entre Universidades e Organizações Mercadológicas, elas possuem algumas diferenças que devem ser consideradas. As universidades estão voltadas para a criação e a disseminação do conhecimento. Algumas metas existem, porém raramente são feitos projetos de pesquisas onde se definem claramente prazos finais. Já com respeito às empresas, há a preocupação com cronogramas, com o cumprimento de metas e outras atividades a curto prazo, no contexto de um ambiente altamente competitivo.

As Universidades e as empresas empregam linguagens distintas; enquanto a primeira se preocupa com a codificação do conhecimento, a segunda está voltada ao conhecimento direcionado à geração de produtos. Por exemplo: hipóteses, modelos e variáveis, termos importantes no idioma dos pesquisadores da Universidade, não possuem a menor importância no vocabulário da maior parte dos representantes das empresas.

Os ambientes de trabalho na universidade e na empresa são bastante diferentes. Para os pesquisadores da universidade, a reputação no meio intelectual é a maior força motivacional, ficando assim o foco de referência situado do lado de fora da organização, em seu grupo de referência profissional.

A Universidade não entende as forças de mercado, as demandas de tempo, e as estruturas de incentivo da empresa. Já na empresa, para a maioria dos gerentes envolvidos nas pesquisas e desenvolvimentos, o superior hierárquico é o referencial crítico. As avaliações de desempenho vêm desta fonte e levam em conta resultados específicos provenientes de sua atuação no trabalho. Da mesma forma, a empresa não entende como tal o trabalho realizado nas Universidades, nem são familiares com os investimentos em recursos humanos e capital físico que precederam sua relação com a Universidade.

Outro ponto crucial, é que os interesses dos pesquisadores da universidade podem mudar, e a universidade os deixa relativamente livres para abandonar determinados projetos e ingressarem em outros mais motivadores.

Os objetivos das duas organizações mercadológicas são bastante diferentes. A maioria das empresas quer aplicações concretas, entrando na relação porque visam o acesso a: procedimentos inovadores, soluções de seus problemas, novo conhecimento científico, novas ferramentas, novas metodologias e novos produtos e serviços. A natureza da pesquisa tecnológica, porém, é complexa, ambígua, e abstrata. Muito do conhecimento gerado pode ser tácito, significando que seus princípios subjacentes são difíceis de identificar e articular. Além disso, provavelmente existirão longos espaços de tempo entre o início do projeto e a criação de produtos. Todas estas características podem criar crises, enganos e dificuldades na transferência do conhecimento.

Já as Universidades trabalham para a obtenção de um produto muito diferente, que pode ser caracterizado a partir de contribuições para o conhecimento, na forma de novos conceitos, modelos, soluções empíricas, técnicas de medidas, e outras contribuições tecnológicas.

2.5 A Pesquisa Científica na Ufla

Esta seção tem como principal objetivo o de situar a Universidade Federal de Lavras na discussão que vem sendo realizada ao longo de todo o trabalho. Para tanto, será apresentado um resumo das principais ações, que estão sendo atualmente tomadas com relação à pesquisa nesta universidade.

As informações desta seção são baseadas em informações colhidas na Pró-Reitoria de Pesquisa (PRP) da Universidade Federal de Lavras (Ufla), através da pesquisa a documentos produzidos pela própria PRP.

A Ufla é uma Instituição Federal de Ensino Superior, localizada na cidade de Lavras, ao sul do estado de Minas Gerais. É uma universidade com 95 anos

de história dedicada à manutenção da alta qualidade do ensino, da pesquisa e da extensão. Atualmente, a Ufla oferece 10 cursos de graduação e 28 cursos de pós-graduação presenciais (destes, 14 de mestrado, 12 de doutorado e 2 de especialização). Diretamente ligadas às atividades de pesquisa da Ufla, estão 302 professores, 2.342 estudantes de graduação e 786 pós-graduandos.

As atividades de pesquisa na Ufla, embora desenvolvidas desde sua fundação no início do século passado, foram modestas até a década de 60, quando houve grande preocupação com a expansão e qualificação do quadro de docentes, o que possibilitou a criação dos cursos de pós-graduação em 1975. Estas ações tiveram grande impacto na pesquisa que, nas últimas duas décadas, experimentou um crescimento acentuado em volume e qualidade, dobrando o número de publicações neste período. Ações e atividades de pesquisa são divulgadas como artigos científicos, conferências, publicações em congressos e boletins técnico-científicos, nos mais diversos temas da ciência e tecnologia.

Os mais de 200 doutores pesquisadores da Ufla, além de inúmeros mestres, pós-graduados, bolsistas de iniciação científica e técnicos de laboratório desenvolvem suas pesquisas em cerca de 60 laboratórios especializados, bem equipados e estruturados para pesquisa científica e ou tecnológica, além de contarem com vários setores temáticos. Desenvolvem, em parcerias com empresas estatais e privadas, inúmeros projetos e programas de cooperação técnico-científico. A Ufla conta atualmente com 65 grupos de pesquisa que desenvolvem em torno de 350 linhas de pesquisa que compõem os projetos isolados e programas especiais. A universidade é bastante competitiva na captação de recursos para a pesquisa nas Agências de Fomento às atividades de Ciência e Tecnologia (C&T) e disponibiliza seus recursos humanos e infra-estrutura para projetos em cooperação e consultorias nas mais diversas áreas de sua atuação. Em seu planejamento estratégico, ações estão sendo implementadas para viabilizar um modelo de gestão eficiente da pesquisa, visando maximizar recursos materiais, humanos e financeiros, de modo a ampliar a pesquisa e aumentar sua aplicabilidade e inserção na sociedade.

As políticas e ações da pesquisa na Ufla evoluíram substancialmente nas últimas décadas, devido, principalmente, a uma maior conscientização sobre a necessidade de melhor equilíbrio entre as atividades de ensino, pesquisa e extensão; a um plano de capacitação de docentes; a expansão e substituição do quadro de docente, que de 1996 a 1999 foi superior a 53%; a consolidação dos programas de pós-graduação e, finalmente, a ampliação e melhoria de infra-estrutura.

O cenário atual exige ações adicionais, tais como, a contratação de novos docentes devido a novas áreas e novas demandas; a constante criação de novos cursos

de graduação e pós-graduação; a inclusão de avanços metodológicos e científicos da C&T mundial e a adaptação ao novo modelo nacional de financiamento e gestão da C&T, que exige maior integração, visibilidade e aplicabilidade.

Desenvolver pesquisa é a grande motivação e incentivo aos docentes devido a valorização pessoal e profissional, a complementaridade da atividade universitária, uma vez que a pesquisa é parte de sua missão, a contribuição à atividade didático-pedagógica, pois evita repasse copiado de informações, a progressão funcional da carreira do docente, o incentivo financeiro, a possibilidades de assessoria/consultoria, como tarefas de extensão, o reforço financeiro para o sistema advindos de auxílios externos, e a facilitação de inserção na comunidade, que é missão social da universidade.

A agenda de trabalho da Ufla atualmente está direcionada à organização interna e a ações extra-campus visando ampliar a dimensão da pesquisa da universidade, tanto em termos de quantidade quanto de qualidade, e promover maior inserção da mesma na sociedade.

A seguir serão apresentados os principais aspectos da Política Institucional para a Pesquisa da Ufla:

- Estabelecer uma estrutura administrativa descentralizada e com ações sistêmicas de controle e gestão das atividades de pesquisa, tendo como instrumento a criação de Comissões de Assessoramento e mecanismos de controle;
- Organizar as atividades de pesquisa a partir de suas bases, adotando-se os "grupos de pesquisa" como unidades de planejamento e de gestão;
- Estabelecer programas institucionais de pesquisa direcionados à ampliação e melhoria da pesquisa desenvolvida na Ufla;
- Promover melhoria na infra-estrutura de pesquisa e viabilizar a implantação de unidades especializadas de apoio à pesquisa;
- Implementar ações sobre propriedade intelectual, transferência de tecnologia e estabelecer mecanismos de proteção do conhecimento gerado na Ufla;
- Promover ações de divulgação sobre a Legislação Nacional de Biossegurança e criar uma Comissão Interna de Biossegurança - CIBio, para viabilizar estudos com organismos geneticamente modificados no Campus;

- Criar mecanismos que facilitem a interação universidade-empresa, em conjunto com a Pró-Reitoria de Extensão e FAEPE, visando ampliar as oportunidades de parcerias externas, no âmbito da C&T (projetos, assessorias, consultorias e serviços);
- Elaborar um plano institucional de C&T com programas retro-alimentados pelos seus resultados (publicações, patentes, produtos e serviços);
- Ampliar as ações do programa de iniciação científica visando maior integração deste às demais atividades acadêmicas da universidade.

A organização da universidade está dividida em Comissões de Assessoramento, cada uma com sua competência bem definida. Dentre elas, existem a comissão responsável pelos Programas de Iniciação Científica, a comissão responsável por Grupos, Projetos e Programas de Pesquisa, a comissão de Infra-Estrutura de Apoio à Pesquisa, dentre outras.

A comissão responsável pela Integração Universidade-Empresa é aquela que deve propor mecanismos e critérios para o estabelecimento de parcerias com empresas; avaliar as propostas de assinatura de convênios de cooperação técnico-científica e outros tipos de atividades de pesquisa em parcerias com empresas; identificar programas na universidade e empresas com potencial para estabelecimento de parcerias com a Ufla; sugerir ações indutoras da interação com empresas para o desenvolvimento de C&T na Ufla; desenvolver estratégias específicas para viabilizar a implantação de núcleos de inovação tecnológica (NIT's) no campus e contribuir para a elaboração e implantação de um programa institucional de parceria universidade-empresas para P&D, em conjunto com a Pró-Reitoria de Extensão.

A Comissão de Assessoramento da Pró-Reitoria de Pesquisa encarregada de avaliar os Grupos, Projetos e Programas de Pesquisa da Ufla iniciou seus trabalhos com vistas a elaboração de um plano institucional de C&T. Após ampla discussão a comissão constatou que os docentes da Ufla têm qualificação e alta produtividade científica, contudo, podem ser mais eficientes no processo de geração e aplicação do conhecimento. Houve grande aumento no número de grupos de pesquisa na Ufla passando de 21 para 59 registrados na última versão 4.0 do Diretório Nacional do CNPq. Considerando as orientações gerais do Diretório, alguns grupos deverão passar por reformulações quanto a composição e definição de linhas de pesquisa, para adequar aos critérios do órgão gestor (CNPq) e a organização da pesquisa na universidade. Há necessidade de melhorar o fluxo de informação das pesquisas geradas na Ufla. É importante que os resultados das pesquisas, sejam difundidos

de modo a efetivamente contribuir para o avanço social e econômico regional e do país.

Há maior possibilidade de sucesso nas propostas de financiamento da pesquisa, quando os projetos individuais ou integrados, são associados a programas institucionais com potencial para efetivamente solucionar problemas da comunidade e do setor produtivo ou que apresentam avanços científicos e inovações tecnológicas. Foram recomendados alguns programas fundamentados na possibilidade de envolvimento de maior número possível de grupos de pesquisa, na relevância econômica e social da atividade, especialmente nas áreas de maior inserção da universidade, levando-se em consideração a tradição, capacitação e localização geográfica da Ufla, na necessidade de que a universidade se envolva mais efetivamente com pesquisas direcionadas à solução de problemas evidentes e avanços sócio-econômico e ambiental, e na demanda real de conhecimento em muitos desses programas. Trata-se, portanto, de uma proposta inicial para direcionar a discussão da programação de pesquisa pela comunidade universitária.

Espera-se que os programas institucionais resultem de ações de pesquisa organizadas e convergentes para objetivos e metas que representem uma política institucional sobre um determinado tema, área ou atividade. Sua organização origina-se dos grupos institucionalmente constituídos, que se organizam a partir de sua competência, linhas de pesquisa e interesses específicos que se aglomeram por temáticas comuns em sub-projetos e projetos independentes ou integrados, até mesmo extra-institucional. Cada programa institucional poderá envolver ou resultar de vários projetos ou sub-projetos desenvolvidos por um ou mais grupos de pesquisa, assim como por um único projeto temático abrangente. Na concepção da proposta, dependendo da especialidade, um grupo ou projeto pode estar inserido em mais de um programa. Casos assim deverão ocorrer com os projetos de grupos ou áreas mais básicas como informática, química e física, dentre outras.

O que se busca é uma organização institucional da pesquisa visando o estabelecimento de políticas gerais para a universidade no âmbito da pesquisa orientada para aspectos de interesse institucional. Isto permitirá melhor planejamento e gestão das atividades em desenvolvimento e de ações futuras. Embora busque-se alguns ajustes de focos da pesquisa atual em alguns segmentos, estes deverão ocorrer, voluntariamente por parte dos pesquisadores quando julgarem necessário ou pertinente. Linhas e projetos inseridos e focados nos programas que refletem os interesses da comunidade científica e da sociedade, referidos como programas institucionais, serão mais coerentes e relevantes e assim mais competitivos na captação de recursos tanto públicos quanto privados, permitindo maior inserção de

nossas ações no ambiente extra-universidade. As linhas de pesquisa ou projetos, que não enquadrarem no universo de um dos programas, continuarão como estão na forma de projetos individuais ou isolados.

A contribuição Científica e Tecnológica da Ufla tem como principais objetivos resgatar os principais resultados da pesquisa na Ufla e fazer uma análise crítica da contribuição e impacto destes para C&T, nas últimas décadas do século XX e difundir e ampliar a participação da Ufla no discurso Científico e Tecnológico nacional. A meta é publicar uma obra científica referencial, no formato de livro multi-autorado, sobre as principais contribuições dos docentes desta universidade nos diferentes tópicos ou temas de sua atuação.

As ações que estão sendo implementadas neste sentido são:

- Levantamento e seleção dos tópicos a serem avaliados mediante consultas aos departamentos, coordenadorias de cursos de pós-graduação e líderes de grupos de pesquisa.
- Organização dos tópicos por áreas afins do conhecimento e definição de coordenadores e editores setoriais da obra científica.
- Definição de autores e co-autores responsáveis pela preparação dos textos temáticos e elaboração de normas específicas e datas para a preparação dos mesmos.
- Preparação do texto sobre uma revisão crítica das contribuições mais relevantes no tema, incluindo uma síntese dos resultados em relação ao estado da arte atual.

O controle das atividades de pesquisa é feito pela Pró-Reitoria de Pesquisa, que verifica se os projetos estão sendo apreciados e aprovados em assembléia departamental, se os departamentos estão estabelecendo um banco de projetos, utilizando um arquivo documental e um banco de dados atualizados, e se o formulário de registro de projeto está sendo preenchido e enviado à Pró-Reitoria para controle institucional. A Pró-Reitoria estabelecerá um banco de dados de projetos de pesquisa e da produção científica a partir dos relatórios de atividades de pessoal docente.

2.6 Plataforma Lattes

Todas as informações desta seção foram extraídas de documentos desenvolvidos pelo Grupo Stela da Universidade Federal de Santa Catarina, que estão disponíveis na própria 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 possibilitar uma gestão estratégica para dar mais qualidade às atividades de fomento do CNPq.

Esses quatro sistemas de informação integrados, articulados com outras bases de dados, localizadas fora da Agência - a base de patentes do Instituto Nacional de Propriedade Industrial (INPI), os bancos de dissertações e teses das universidades - constituem a Plataforma Lattes.

2.6.1 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 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.

Além disso, as instituições podem desenvolver suas próprias rotinas para importação dos dados curriculares para suas bases corporativas, uma vez que os dados extraídos estejam armazenados num formato aberto e documentado, em arquivos XML. As extrações do banco de currículos do CNPq são feitas em lotes e podem ser configuradas de acordo com o interesse e com as permissões de cada usuário do Lattes Extrator.

O funcionamento do Lattes Extrator depende da formação, no CNPq, de um Data Warehouse com todos os currículos em XML. Tal ambiente deverá ser atualizado constantemente, caso contrário a extração poderá fornecer dados desatualizados às instituições usuárias. Tendo em vista o enorme aumento de chegada de currículos Lattes ao CNPq, em função do lançamento da versão 5.0 do Diretório de Grupos de Pesquisa e dos editais de fomento, optou-se por disponibilizar o Lattes Extrator em fase experimental.

Atualmente, estão sendo oferecidos para extração os currículos recebidos pelo CNPq até o dia 15 de julho de 2002, prazo de encerramento da coleta dos dados para o Censo 2002 dos Grupos de Pesquisa. De julho até novembro de 2002 foram inseridos aproximadamente 25 mil novos currículos na Plataforma Lattes e cerca de 40 mil foram atualizados. Tão logo seja possível atualizar o Data Warehouse do CNPq com os XMLs dos currículos recebidos após o encerramento do Censo 2002, estaremos disponibilizando-os para extração.

O acesso ao Lattes Extrator depende do credenciamento prévio, junto ao CNPq, da instituição interessada. Para tal, a instituição deve seguir os seguintes procedimentos: a instituição deve encaminhar carta ou comunicação oficial de seu representante máximo para o presidente do CNPq solicitando o credenciamento da instituição para uso do sistema Lattes Extrator e justificando o seu uso. Deve também preencher o Termo de Compromisso referente ao uso do sistema Lattes Extrator e encaminhá-lo ao CNPq. Além da carta de comunicação e do Termo de Compromisso preenchido, a instituição deverá reunir os seguintes documentos: cópia da portaria de nomeação do representante da instituição que assinará o Termo de Compromisso; e cópia do estatuto da instituição. A área técnica do CNPq verifica os documentos recebidos e, após parecer afirmativo, encaminha a senha de acesso ao Sistema Lattes Extrator para o representante da instituição que assinou o Termo de Compromisso.

Capítulo 3

Metodologia

3.1 Tipos de Pesquisa

Para o desenvolvimento deste trabalho, foram utilizadas as Pesquisas Bibliográfica e Documental e a Metodologia de Estudo de Caso.

A Pesquisa Bibliográfica deu base para a aquisição de conhecimento acerca dos temas envolvidos no projeto, como, por exemplo, Gestão do Conhecimento, mecanismos de Descoberta de Conhecimento em Bancos de Dados e técnicas para a construção do sistema de mineração de dados. Envolveu, basicamente, consultas a livros de referência e a teses e artigos científicos.

A pesquisa documental foi realizada nos documentos referentes à Pesquisa Científica na Ufla, gerados a partir da Plataforma Lattes e trouxeram informações úteis para as análises, as comparações e para o desenvolvimento da ferramenta de *Data Mining*.

O Estudo de Caso é a própria Universidade Federal de Lavras, mais especificamente os setores envolvidos com o desenvolvimento de pesquisa científica. O estudo foi baseado em análises dos currículos dos docentes e de alguns discentes da Ufla. Também foram feitas entrevistas com pessoas-chave envolvidas diretamente com a gestão da Ufla para coletar informações e/ou dados que pudessem auxiliar na criação e análise do processo de Descoberta de Conhecimento. Os resultados destas entrevistas foram utilizados na etapa de filtragem dos dados.

3.2 Procedimento Metodológico

Foi realizada, inicialmente, uma pesquisa bibliográfica dos assuntos discutidos nas seções anteriores para adquirir um embasamento teórico. Foi necessário conhecer todo o campo da Descoberta de Conhecimento em Bancos de Dados e *Data Mining*: processos, modelos, técnicas utilizadas, entre outros.

Além disso, foi necessário, também, conhecer os problemas enfrentados pelos gestores da Universidade Federal de Lavras, no que diz respeito às Pesquisas Científicas. Conhecer suas prioridades, metas, mecanismos e cultura, de um modo geral, foi essencial para a realização deste trabalho.

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 3.1, 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 3.1: 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 subdivi-

dida em duas fases: a primeira consistindo basicamente de cruzamento de consultas SQL ¹ 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.

3.3 Desenvolvimento

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

¹*Structured Query Language* - linguagem de Banco de Dados.

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. Esta etapa será melhor explicada ainda nesta seção.
6. *Data Mining*: A etapa consistiu, basicamente, da primeira fase do *Data Mining*, com elaboração, cruzamento e comparação de consultas SQL.
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.

A Figura 3.2, a seguir, mostra a identificação dessas etapas realizadas com as etapas pré-definidas do processo de DCBD:

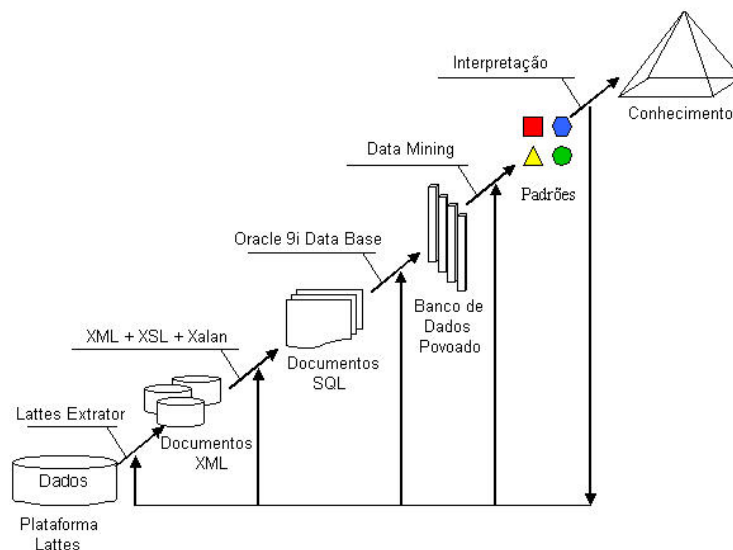


Figura 3.2: Execução das etapas do processo de DCBC neste trabalho.

Na etapa de codificação dos dados, 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. Ambos sistemas foram utilizados em uma máquina Duron, de 1GHz, com 256 MB de memória RAM, que se encontra no Departamento de Ciência da Computação, na Ufla.

Foi escolhido o software da Oracle pelo fato de ele apresentar facilidade de uso, potência e uma boa relação custo-desempenho. Ele também contém um conjunto totalmente integrado de ferramentas de gerenciamento simples de usar, além de recursos completos de distribuição, replicação e, o mais importante para este trabalho, a ferramenta de *Data Mining*.

Antes de iniciar a descrição do desenvolvimento dos sistemas, é necessário abordar os conceitos de XML e XSL.

Extensible Markup Language (XML)

A *Extensible Markup Language* 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, ao contrário da HTML que possibilita apenas que os documentos sejam lidos pelas pessoas, e não otimizados para tratamento por computadores.

Segundo Deitel (2003), a XML é uma linguagem de marcação que descreve dados de praticamente qualquer tipo de forma estruturada hierarquicamente. Diferentemente da HTML, que possui um conjunto fixo de marcações, a XML permite que a descrição dos dados de forma mais precisa através da criação de novas marcas.

A XML define a estrutura de um documento, oferecendo uma abordagem padrão para descrição, captura, processamento e publicação de informações. As características essenciais da XML, segundo Deitel (2003), são: separação de dados, separação do conteúdo e sua apresentação. A ausência de instruções de formatação facilita a realização de análise sintática. Isso torna a XML uma estrutura de referência que pode ser usada para o intercâmbio de dados. Reconhecendo esse fato, os desenvolvedores de software em todo o mundo estão integrando a XML

a seus aplicativos, de forma a obter ganhos de funcionalidade e interoperabilidade na Web.

Contudo, a XML não está limitada à aplicações Web. Ela vem sendo cada vez mais utilizada em bancos de dados, pois a natureza estruturada, mas não formatada, de um documento XML, permite que ele seja manipulado por aplicativos de bancos de dados.

Como documentos XML são estruturados hierarquicamente, eles podem ser representados por uma estrutura de árvore. Os documentos XML extraídos da Plataforma Lattes para este trabalho são constituídos de cinco partes principais: Dados Gerais, Produção Bibliográfica, Produção Técnica, Outra Produção e Dados Complementares.

A XML não se limita a descrever somente dados textuais, também pode descrever imagens, gráficos vetoriais, animações ou qualquer outro tipo de dado para o qual seja estendida [Deitel (2003)].

Atualmente, diversas empresas de porte estão fazendo uso do XML para situações específicas. Bancos podem usar os recursos da estruturação de dados para distribuir o processamento dos mesmos, racionalizando o uso dos recursos de processamento envolvidos; empresas podem acessar dados de parceiros, conforme comentamos acima, de forma a diminuir os custos com a transmissão e compartilhamento de dados; sites podem oferecer parte de seu conteúdo para ser inserido em outros sites, aumentando o número de visitas e etc.

Apesar de todas essas vantagens, para a Descoberta de Conhecimento 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.

Extensible Stylesheet Language (XSL)

A *Extensible Stylesheet Language* é uma linguagem que permite transformar documentos XML em diversos outros formatos, como HTML, texto simples ou qualquer outro documento baseado em texto.

O documento de formatação XSL é chamado de XSLT. A XSL especifica a estilização de um documento XML utilizando a XSLT para descrever como o documento é transformado. Uma transformação expressa em XSLT descreve regras para transformar uma árvore fonte em uma árvore resultado. A transformação é obtida associando-se padrões com gabaritos. Um padrão é confrontado com elementos na árvore fonte e um gabarito é instanciado para criar parte da árvore resultado. A árvore resultado é separada da árvore fonte. A estrutura da árvore resultado

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<CURRICULO-VITAE>
  <DADOS-GERAIS NOME-COMPLETO="Olinda Nogueira Paes Cardoso"
    NOME-EM-CITACOES-BIBLIOGRAFICAS="CARDOSO, O. N. P."
    NACIONALIDADE="B" PAIS-DE-NASCIMENTO="Brasil" UF-NASCIMENTO="BA"
    CIDADE-NASCIMENTO="Valença" DATA-NASCIMENTO="06021972"
    SEXO="FEMININO" NOME-DO-PAI="Edson Edmundo Barreto Paes Cardoso"
    NOME-DA-MAE="Iracema Müller Nogueira"
    PERMISSAO-DE-DIVULGACAO="SIM" OUTRAS-INFORMACOES-RELEVANTES=" ">

```

Figura 3.3: Parte do código XML utilizado na conversão.

pode ser completamente diferente da estrutura da árvore fonte. Na construção da árvore resultado, os elementos da árvore fonte podem ser filtrados e reordenados, podendo-se ainda adicionar estrutura arbitrária [W3C (2004)].

Em outras palavras, a XSL é uma linguagem para criação de folhas de estilo para documentos XML. Por possuir essa funcionalidade, a linguagem XSL foi utilizada, juntamente com o software Xalan, citado anteriormente, para criar uma folha de estilo para o currículo extraído da Plataforma Lates. Essa folha de estilo identifica cada nó da “árvore” XML, ou seja, cada itag do documento XML, extraindo seus valores e estruturando-os num código SQL que, depois, seria executado no SGBD da Oracle, para povoar o Banco de Dados.

As Figuras 3.3, 3.4 e 3.5, ilustram esta explicação. A Figura 3.3 mostra uma parte do arquivo XML, extraída do currículo da Professora Olinda Nogueira Paes Cardoso, orientadora deste projeto. O software para a transformação de documentos XML em outros formatos (Xalan) trabalha associando o documento XML (.xml) a ser transformado a outro documento, que é a folha de estilo específica desse XML e encontra-se no formato XSL (.xsl), como mostrado na Figura 3.4. Esse arquivo XSL deve conter todas as instruções necessárias para a conversão, como identificação das tags, dos atributos e dos valores destes. É apenas este o papel do software auxiliar utilizado neste trabalho: comparar os dois documentos, fazer uma associação entre as estruturas e identificar os valores. O arquivo de saída é um documento no formato SQL.

As estruturas entre os símbolos < e > são as *tags* do documento XML, como, por exemplo, <CURRICULO-VITAE>. Os atributos são as palavras que vêm logo depois do nome da *tag*, como, por exemplo, o atributo “nome completo” em <DADOS-GERAIS NOME-COMPLETO = "Olinda Nogueira Paes Cardoso">. O valor do atributo é tudo o que vem depois do “=” e entre “aspas”. No caso deste

último exemplo, o valor do atributo “nome-completo” seria “Olinda Nogueira Paes Cardoso”.

O código XSL criado para a extração das informações desse currículo seria como na Figura 3.4:

A parte do código `<xsl:template match = "DADOS-GERAIS">` é a estrutura XSL que identifica a *tag* DADOS-GERAIS do documento XML. Encontrada essa *tag*, é inserido o código SQL e, dentro deste, a estrutura “`<xsl:value-of select = @..."` extrai o valor do atributo especificado após o “@” que deverá ser inserido no Banco de Dados. O arquivo de saída é como o da Figura 3.5.

Antes da inserção dos dados, foi necessária a modelagem e a criação do Banco de Dados. Para tanto, utilizou-se o SGBD Oracle, versão 9i, como citado anteriormente. A modelagem deu origem às tabelas ilustradas no Apêndice A.

Depois da modelagem, criação e ajuste do Banco de Dados, foram executados, no SGBD Oracle 9i, todos os arquivos SQL gerados a partir da transformação dos documentos XML para inserir todos os dados dos currículos extraídos da Plataforma Lattes, totalizando 575 documentos (que incluíam, além dos currículos dos professores, currículos de ex-alunos, mestrandos e doutorandos da Ufla).

Iniciou-se, após esta etapa, a fase de filtragem dos dados. Foram excluídos dados repetidos, campos das tabelas que estavam em branco e informações erradas, inseridas pelo próprio usuário na hora de cadastrar os dados no currículo lattes. Exemplo de alguns campos que apareceram em branco em alguns documentos: título da dissertação de mestrado, área do conhecimento do artigo publicado, setor de atividade de um trabalho em evento e até mesmo o nome da instituição com a qual possui, ou possuiu, vínculo profissional.

Filtrados os dados, passou-se à primeira fase da etapa de mineração dos dados ou *Data Mining*. Primeiramente, 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 ferramenta de *Data Mining* na fase 2 desta etapa. Algumas dessas consultas foram:

- Quantas pessoas cadastradas ainda possuem vínculo profissional com a Ufla?
- Que tipo de vínculos profissionais são permitidos na Ufla?
- Quais as áreas de maior atuação profissional?
- Quantas áreas e sub-áreas do conhecimento diferentes foram cadastradas?
- Existem produções científicas de várias pessoas na mesma especialidade?

```

<?xml version = "1.0" encoding="ISO-8859-1"?>
<xsl:stylesheet xmlns:xsl =
    "http://www.w3.org/1999/XSL/Transform" version = "1.0">

<xsl:template match = "CURRICULO-VITAE">
    <xsl:apply-templates/>
</xsl:template>

<xsl:template match = "DADOS-GERAIS">

DECLARE PESSOA NUMBER;
    CODPOSGRAD NUMBER;
    CODATUACAO NUMBER;
    CODAREA NUMBER;
    CODPD NUMBER;
    CODENSINO NUMBER;
    CODTREINAMENTO NUMBER;
    CODTRAB NUMBER;
    CODARTIGO NUMBER;
    CODPRODUCAO NUMBER;
    CODTRABTEC NUMBER;
    CODORIENTACAO NUMBER;
    CODPARTICIPACAO NUMBER;

BEGIN
SELECT COUNT (COD_PESSOA)+1 INTO PESSOA FROM DADOS_GERAIS;
INSERT INTO DADOS_GERAIS(COD_PESSOA, NOME, NOME_CITACOES,
NACIONALIDADE, PAIS_NASC, UF_NASC, CIDADE_NASC, DATA_NASC,
SEXO, NOME_PAI, NOME_MAE, FLAG_DIVULGACAO, OUTRAS_INF)
values (peessoa,
'<xsl:value-of select = "@NOME-COMPLETO" />',
'<xsl:value-of select="@NOME-EM-CITACOES-BIBLIOGRAFICAS" />',
'<xsl:value-of select = "@NACIONALIDADE" />',
'<xsl:value-of select = "@PAIS-DE-NASCIMENTO" />',
'<xsl:value-of select = "@UF-NASCIMENTO" />',
'<xsl:value-of select = "@CIDADE-NASCIMENTO" />',
'<xsl:value-of select = "@DATA-NASCIMENTO" />',
'<xsl:value-of select = "@SEXO" />',
'<xsl:value-of select = "@NOME-DO-PAI" />',
'<xsl:value-of select = "@NOME-DA-MAE" />',
'<xsl:value-of select = "@PERMISSAO-DE-DIVULGACAO" />',
'<xsl:value-of select="@OUTRAS-INFORMACOES-RELEVANTES" />'
);
<xsl:apply-templates select="ENDERECO" />
<xsl:apply-templates select="FORMACAO-ACADEMICA-TITULACAO" />
<xsl:apply-templates select="ATUACOES-PROFISSIONAIS" />
</xsl:template>

```

Figura 3.4: Parte do código XSL utilizado na conversão.

```

DECLARE PESSOA NUMBER;
        CODPOSGRAD NUMBER;
        CODATUACAO NUMBER;
        CODAREA NUMBER;
        CODPD NUMBER;
        CODENSINO NUMBER;
        CODTREINAMENTO NUMBER;
        CODTRAB NUMBER;
        CODARTIGO NUMBER;
        CODPRODUCAO NUMBER;
        CODTRABTEC NUMBER;
        CODORIENTACAO NUMBER;
        CODPARTICIPACAO NUMBER;

BEGIN
SELECT COUNT (COD_PESSOA)+1 INTO PESSOA FROM DADOS_GERAIS;
INSERT INTO DADOS_GERAIS(COD_PESSOA, NOME, NOME_CITACOES,
NACIONALIDADE, PAIS_NASC, UF_NASC, CIDADE_NASC, DATA_NASC,
SEXO, NOME_PAI, NOME_MAE, FLAG_DIVULGACAO, OUTRAS_INF)
values (peessoa,
        'Olinda Nogueira Paes Cardoso',
        'CARDOSO, O. N. P.',
        'B',
        'Brasil',
        'BA',
        'Valença',
        '06021972',
        'FEMININO',
        'Edson Edmundo Barreto Paes Cardoso',
        'Iracema Müller Nogueira',
        'SIM',
        ''
);

```

Figura 3.5: Parte do script SQL gerado a partir da conversão.

- Quais são as linhas de pesquisa existentes, hoje, na Ufla?
- A que órgãos estão ligadas as atividades de pesquisa?
- Quantos artigos já foram publicados por pessoas ligadas à Ufla?
- Quantos artigos cada pessoa publicou?
- Que área do conhecimento possui mais artigos publicados?
- Quais os locais de mais publicação de artigos da Ufla?
- Cada atividade de pesquisa de uma pessoa se encaixa em quantas linhas de pesquisa?
- Quantas das atuações profissionais de cada pessoa são atividades de ensino?
- Quantas atividades de ensino por semestre cada pessoa realizou?
- Quantas atividades de ensino por curso e por semestre?
- Qual o número máximo e mínimo de atuações profissionais de 1 pessoa por semestre?
- Quantos trabalhos em eventos foram realizados em cada área do conhecimento?
- Quantos artigos foram publicados em cada área do conhecimento?
- O mesmo artigo foi cadastrado por dois autores diferentes?
- O mesmo trabalho foi cadastrado por dois autores diferentes?

Essas e muitas outras consultas foram feitas utilizando a interface de linha de comando do SGBD Oracle. Algumas dessas linhas de comando utilizadas para fazer as consultas, estão ilustradas na Figura 3.6.

Ainda poderão ser feitas, após a realização da fase 2 do *Data Mining*, outras consultas, porém mais complexas, como: (1) Qual a implicação de uma atividade de direção para outras atuações por semestre, como atividades de pesquisa, trabalhos em eventos, serviços técnicos, número de orientações e de disciplinas ministradas? (2) A quantidade de artigos publicados por semestre é igual à quantidade de orientações? (3) O número de disciplinas ministradas por um professor

```

1) Quantidade de atividades de direção que cada pessoa realizou.
SQL> select count(cod_atuacao), cod_pessoa from atuacoes
      where tipo = 'AD' group by cod_pessoa;

2) Quantidade de atuações com atividades de ensino.
SQL> select count(*), atuacoes.cod_pessoa, nome
      from atuacoes, dados_gerais
      where dados_gerais.cod_pessoa=atuacoes.cod_pessoa and
            atuacoes.cod_atuacao in (select atividade_ensino.cod_atuacao
            from atividade_ensino)
      group by atuacoes.cod_pessoa, nome
      order by 1 desc;

3) Número de artigos publicados por cada pessoa ligada à Ufla.
SQL> select distinct dados_gerais.nome, artigos_publicados.cod
      from dados_gerais, artigos_publicados
      where dados_gerais.cod_pessoa = artigos_publicados.cod_pessoa

4) f Pessoas que não têm atuações cadastradas.
SQL> select nome from dados_gerais
      where cod_pessoa not in (select cod_pessoa from atuacoes)
      order by 1;

5) Quantidade de atuações com pesquisa.
SQL> select count(*), atuacoes.cod_pessoa, nome
      from atuacoes, dados_gerais
      where dados_gerais.cod_pessoa=atuacoes.cod_pessoa and
            atuacoes.cod_atuacao in (select atividade_pesquisa.cod_atuacao
            from atividade_pesquisa)
      group by atuacoes.cod_pessoa, nome
      order by 1 desc;

```

Figura 3.6: Exemplo do código de algumas consultas SQL feitas ao Banco de Dados deste trabalho.

influencia na quantidade de artigos que ele publica? (4) Existe alguma relação entre a publicação de um artigo no exterior e o fato do autor ter sido pós-graduado no mesmo? (5) Qual a probabilidade de um professor, que já realizou alguma atividade de direção, vir a se candidatar a diretor em um outro semestre?

Esses são apenas alguns exemplos do vasto campo de conhecimento que pode ser descoberto através da utilização do sistema criado neste trabalho e através da atualização do processo de DCBD iniciado também neste. Outras descobertas e a execução da fase 2 do *Data Mining* compreendem trabalhos futuros.

Capítulo 4

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 Lattes 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 4.1 demonstra essa variedade.

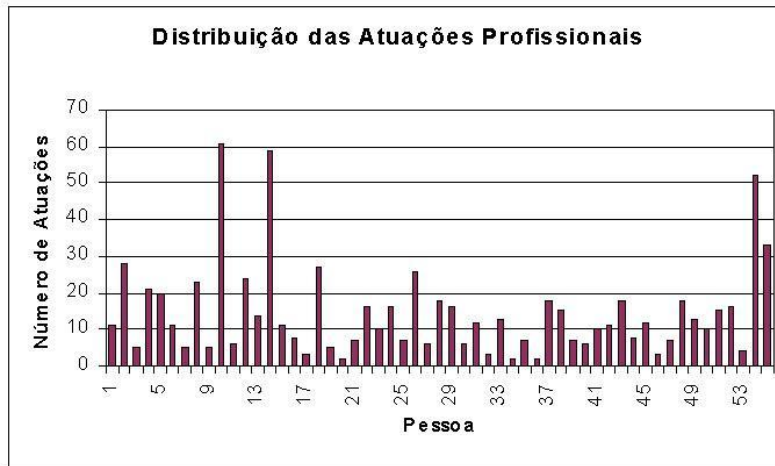


Figura 4.1: Número de atuações profissionais por pessoa.

O que se pôde observar das atividades cadastradas pelos 10% das pessoas, é que 6% delas realizaram entre 1 e 16 atividades de ensino, como mostra a Figura 4.2. Essas atividades de ensino incluíam ensino médio, graduação, pós-graduação e especialização.

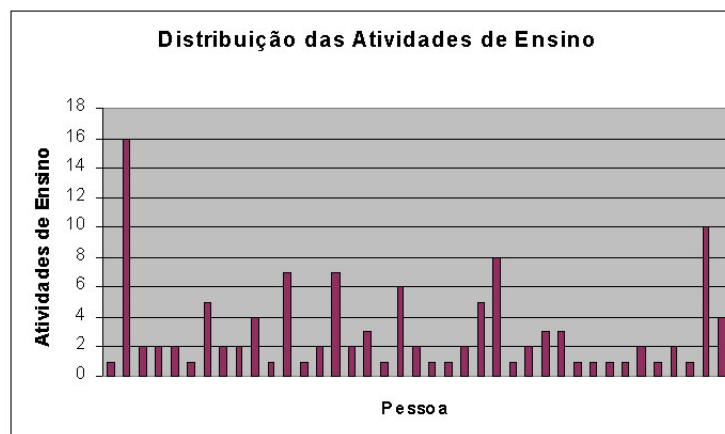


Figura 4.2: Número de atividades de ensino por pessoa.

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, distribuição ilustrada pelo gráfico da Figura 4.3.

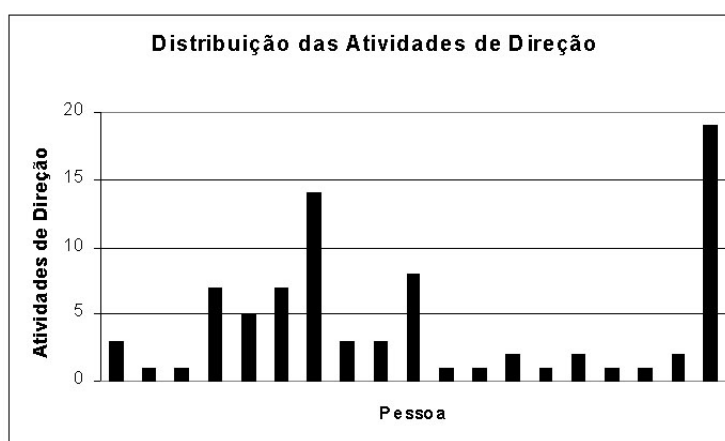


Figura 4.3: Número de atividades de direção por pessoa.

No que se refere às atividades de pesquisa, 7% realizaram entre 1 e 6 atividades de pesquisa, como mostra o gráfico da Figura 4.4. Cada uma dessas atividades compreendia, na maioria das vezes, de 1 a 3 linhas de pesquisa, mas algumas possuíam até 9 linhas.

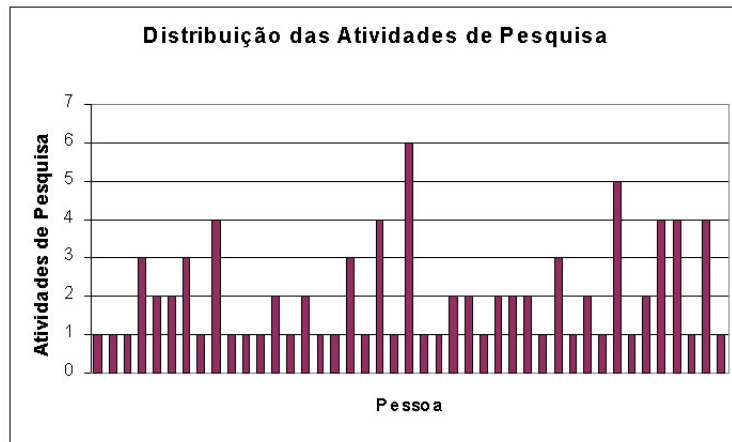


Figura 4.4: Número de atividades de pesquisa por pessoa.

Além das atividades citadas anteriormente, ainda foram cadastradas 2,6% de atividades de extensão, como consultorias, estágios, cursos, etc. Os 10% de pessoas que cadastraram alguma atividade, realizaram entre 1 e 6 atividades de extensão, como mostra o gráfico da Figura 4.5.

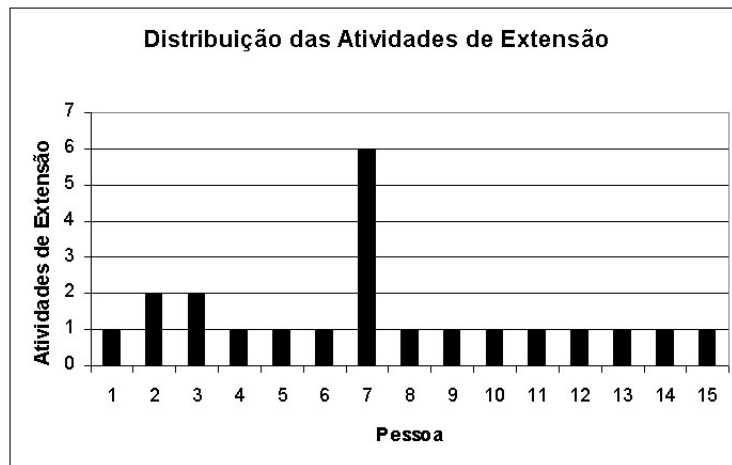


Figura 4.5: Número de atividades de extensão por pessoa.

Além dessas, 2,4% realizaram entre 1 e 32 serviços técnicos e menos de 2% ministraram algum treinamento, distribuições ilustradas pelos gráficos das Figuras 4.6 e 4.7, respectivamente. Entre as atividades de serviço técnico, encontravam-se: assistência técnica, prestação de serviços, entre outros.

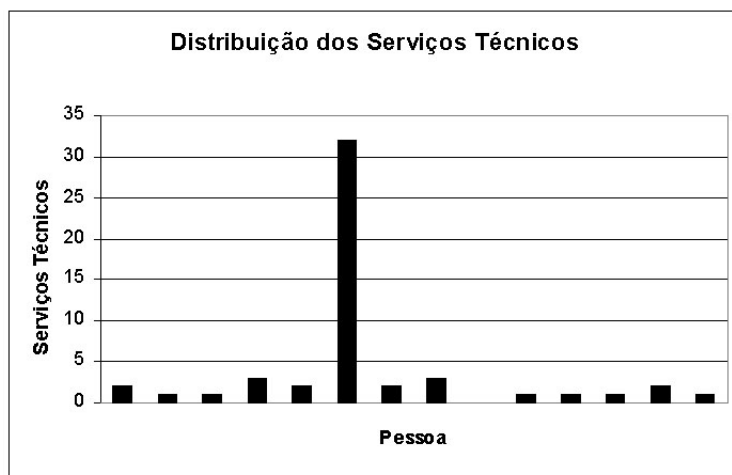


Figura 4.6: Número de serviços técnicos por pessoa.



Figura 4.7: Número de treinamentos ministrados por pessoa.

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, como mostra o gráfico da Figura 4.8.

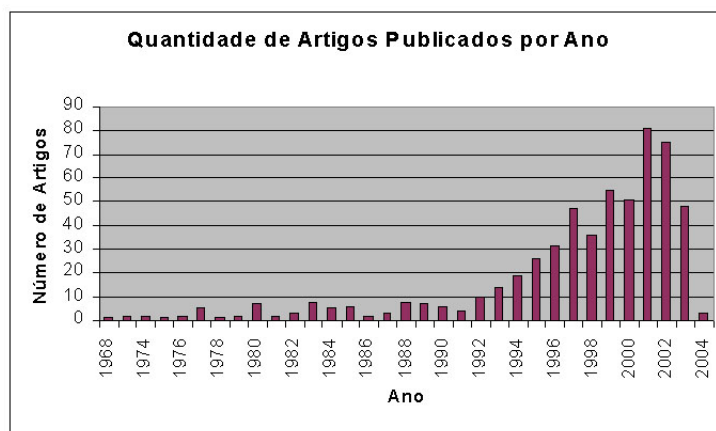


Figura 4.8: Número de artigos publicados por ano.

Dentre esses artigos, 6,4% foram publicados no exterior e a maioria possui de 3 a 5 autores, com alguns possuindo até 8 autores, como ilustra o gráfico da Figura 4.9.

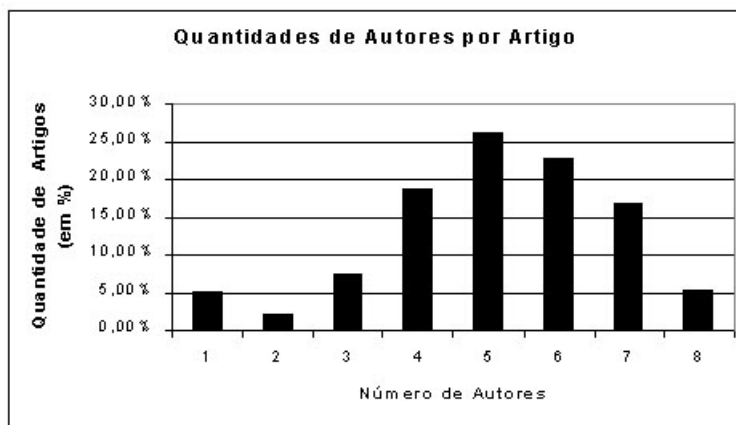


Figura 4.9: Número de autores por artigo.

A Figura 4.10 mostra um gráfico com a distribuição dos artigos por área do conhecimento. De acordo com o site oficial do CNPq ¹, 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.
- Ciências Humanas: Filosofia, Sociologia, Arqueologia, História, Geografia, Psicologia, Educação, Teologia, entre outros.

¹Site <http://www.cnpq.br>, visitado em 26 de junho de 2004.

- 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.

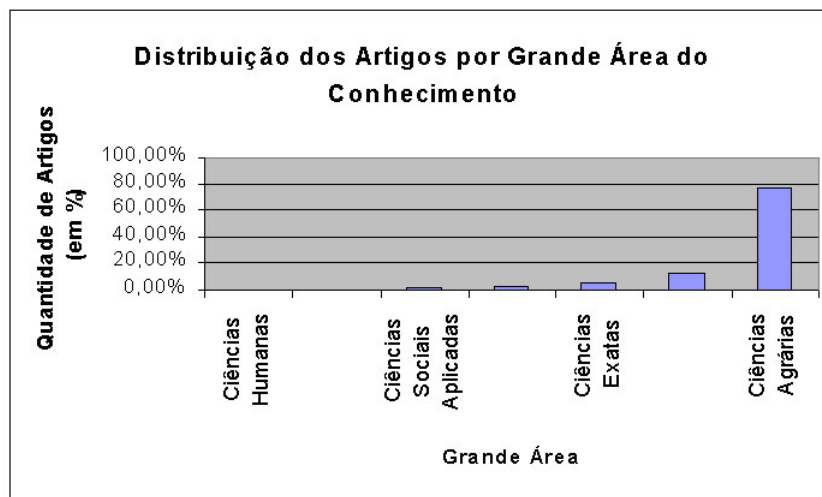


Figura 4.10: Número de artigos publicados por grande área do conhecimento.

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.

Capítulo 5

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 Conhe-

cimento 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 Bibliográficas

- [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.

Apêndice A

Tabelas do BD Lattes

Apêndice B

Código Fonte XSL

```
<?xml version = "1.0"encoding="ISO-8859-1"?>  
<xsl:stylesheet xmlns:xsl = "http://www.w3.org/1999/XSL/Transform"version =  
"1.0»  
<xsl:template match = "CURRICULO-VITAE»  
<xsl:apply-templates/>  
</xsl:template>
```

```
<xsl:template match = "DADOS-GERAIS»  
DECLARE PESSOA NUMBER;  
CODPOSGRAD NUMBER;  
CODATUACAO NUMBER;  
CODAREA NUMBER;  
CODPD NUMBER;  
CODENSINO NUMBER;  
CODTREINAMENTO NUMBER;  
CODTRAB NUMBER;  
CODARTIGO NUMBER;  
CODPRODUCAO NUMBER;  
CODTRABTEC NUMBER;  
CODORIENTACAO NUMBER;  
CODPARTICIPACAO NUMBER;  
BEGIN  
SELECT COUNT (COD_PESSOA)+1 INTO PESSOA FROM DADOS_GERAIS;  
INSERT INTO DADOS_GERAIS(COD_PESSOA, NOME, NOME_CITACOES,
```

NACIONALIDADE, PAIS_NASC, UF_NASC, CIDADE_NASC, DATA_NASC,
SEXO, NOME_PAI, NOME_MAE, FLAG_DIVULGACAO, OUTRAS_INF)
values (pessoa,

```
'<xsl:value-of select = "@NOME-COMPLETO"/>',  
'<xsl:value-of select = "@NOME-EM-CITACOES-BIBLIOGRAFICAS"/>',  
'<xsl:value-of select = "@NACIONALIDADE"/>',  
'<xsl:value-of select = "@PAIS-DE-NASCIMENTO"/>',  
'<xsl:value-of select = "@UF-NASCIMENTO"/>',  
'<xsl:value-of select = "@CIDADE-NASCIMENTO"/>',  
'<xsl:value-of select = "@DATA-NASCIMENTO"/>',  
'<xsl:value-of select = "@SEXO"/>',  
'<xsl:value-of select = "@NOME-DO-PAI"/>',  
'<xsl:value-of select = "@NOME-DA-MAE"/>',  
'<xsl:value-of select = "@PERMISSAO-DE-DIVULGACAO"/>',  
'<xsl:value-of select = "@OUTRAS-INFORMACOES-RELEVANTES"/>'  
);  
<xsl:apply-templates select = "ENDERECO"/>  
<xsl:apply-templates select = "FORMACAO-ACADEMICA-TITULACAO"/>  
<xsl:apply-templates select = "ATUACOES-PROFISSIONAIS"/>  
</xsl:template>
```

```
<xsl:template match = "ENDERECO»  
<xsl:apply-templates/>  
</xsl:template>  
<xsl:template match = "ENDERECO-PROFISSIONAL»  
INSERT INTO END_PROF  
(COD_PESSOA, NOME_INST, NOME_UNID, NOME_ORGAO, PAIS, UF,  
COMPLEMENTO, BAIRRO, CIDADE, CXPOSTAL, CEP, DDD, TELEFONE,  
RAMAL, FAX, EMAIL, HOME_PAGE)  
VALUES (pessoa,  
'<xsl:value-of select = "@NOME-INSTITUICAO-EMPRESA"/>',  
'<xsl:value-of select = "@NOME-UNIDADE"/>',  
'<xsl:value-of select = "@NOME-ORGAO"/>',  
'<xsl:value-of select = "@PAIS"/>',  
'<xsl:value-of select = "@UF"/>',  
'<xsl:value-of select = "@LOGRADOURO-COMPLEMENTO"/>',  
'<xsl:value-of select = "@BAIRRO"/>',
```

```
'<xsl:value-of select = "@CIDADE"/>',
'<xsl:value-of select = "@CAIXA-POSTAL"/>',
'<xsl:value-of select = "@CEP"/>',
'<xsl:value-of select = "@DDD"/>',
'<xsl:value-of select = "@TELEFONE"/>',
'<xsl:value-of select = "@RAMAL"/>',
'<xsl:value-of select = "@FAX"/>',
'<xsl:value-of select = "@E-MAIL"/>',
'<xsl:value-of select = "@HOME-PAGE"/>'
);
</xsl:template>
```

```
<xsl:template match = "FORMACAO-ACADEMICA-TITULACAO»
<xsl:apply-templates/>
</xsl:template>
```

```
<xsl:template match = "GRADUACAO»
INSERT INTO GRADUACAO (COD_PESSOA, SEQ_FORM, NIVEL,
TITULO_CONCLUSAO, NOME_ORIENTADOR, NOME_INST,
NOME_CURSO, STATUS_CURSO, ANO_INICIO,
ANO_CONCLUSAO, FLAG_BOLSA, NOME_AGENCIA)
VALUES (pessoa,
'<xsl:value-of select = "@SEQUENCIA-FORMACAO"/>',
'<xsl:value-of select = "@NIVEL"/>',
'<xsl:value-of select="@TITULO-DO-TRABALHO-CONCLUSAO-DE-CURSO"/>',
'<xsl:value-of select = "@NOME-DO-ORIENTADOR"/>',
'<xsl:value-of select = "@NOME-INSTITUICAO"/>',
'<xsl:value-of select = "@NOME-CURSO"/>',
'<xsl:value-of select = "@STATUS-DO-CURSO"/>',
'<xsl:value-of select = "@ANO-DE-INICIO"/>',
'<xsl:value-of select = "@ANO-DE-CONCLUSAO"/>',
'<xsl:value-of select = "@FLAG-BOLSA"/>',
'<xsl:value-of select = "@NOME-AGENCIA"/>'
);
</xsl:template>
```

```
<xsl:template match = "MESTRADO»
```

```

SELECT COUNT (COD_POSGRAD)+1 into CODPOSGRAD FROM
POS_GRADUACAO;
INSERT INTO POS_GRADUACAO (COD_POSGRAD, COD_TIPO,
COD_PESSOA, SEQ_FORM, NIVEL, NOME_INST, NOME_CURSO,
STATUS_CURSO, ANO_INICIO, ANO_CONCLUSAO, FLAG_BOLSA,
NOME_AGENCIA, ANO_OBTENCAO_TITULO, TITULO DISSERTACAO,
NOME_ORIENTADOR)
VALUES (codposgrad,
'1',
pessoa,
'<xsl:value-of select = "@SEQUENCIA-FORMACAO"/>',
'<xsl:value-of select = "@NIVEL"/>',
'<xsl:value-of select = "@NOME-INSTITUICAO"/>',
'<xsl:value-of select = "@NOME-CURSO"/>',
'<xsl:value-of select = "@STATUS-DO-CURSO"/>',
'<xsl:value-of select = "@ANO-DE-INICIO"/>',
'<xsl:value-of select = "@ANO-DE-CONCLUSAO"/>',
'<xsl:value-of select = "@FLAG-BOLSA"/>',
'<xsl:value-of select = "@NOME-AGENCIA"/>',
'<xsl:value-of select = "@ANO-DE-OBTENCAO-DO-TITULO"/>',
'<xsl:value-of select = "@TITULO-DA-DISSERTACAO-TESE"/>',
'<xsl:value-of select = "@NOME-COMPLETO-DO-ORIENTADOR"/>'
);
INSERT INTO PALAVRAS_CHAVE_POSGRAD(COD_POSGRAD, PALAVRA1,
PALAVRA2, PALAVRA3, PALAVRA4, PALAVRA5, PALAVRA6)
VALUES (codposgrad,
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-1"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-2"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-3"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-4"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-5"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-6"/>'
);

<xsl:for-each select = "AREAS-DO-CONHECIMENTO/*">
SELECT COUNT(COD_AREA)+1 INTO CODAREA FROM AREA_POSGRAD;
INSERT INTO AREA_POSGRAD (COD_POSGRAD, COD_AREA,

```

```

GRANDE_AREA, AREA, SUB_AREA, ESPECIALIDADE)
VALUES (codposgrad,
codarea,
'<xsl:value-of select="@NOME-GRANDE-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select="@NOME-DA-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select="@NOME-DA-SUB-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select="@NOME-DA-ESPECIALIDADE"/>'
);
</xsl:for-each>

```

```

INSERT INTO SETOR_ATIVIDADE_POSGRAD (COD_POSGRAD, SETOR1,
SETOR2, SETOR3)
VALUES (codposgrad,
'<xsl:value-of select="SETORES-DE-ATIVIDADE/@SETOR-DE-
ATIVIDADE-1"/>',
'<xsl:value-of select="SETORES-DE-ATIVIDADE/@SETOR-DE-
ATIVIDADE-2"/>',
'<xsl:value-of select="SETORES-DE-ATIVIDADE/@SETOR-DE-
ATIVIDADE-3"/>'
);
</xsl:template>

```

```

<xsl:template match = "ATUACOES-PROFISSIONAIS»
<xsl:apply-templates/>
</xsl:template>

```

```

<xsl:template match = "ATUACAO-PROFISSIONAL»
SELECT COUNT(COD_ATUACAO)+1 INTO CODATUACAO FROM
ATUACOES; INSERT INTO ATUACOES(COD_ATUACAO, COD_PESSOA,
TIPO)
VALUES (codatuacao,
pessoa,
'AP');
INSERT INTO ATUACAO_PROF (COD_ATUACAO, NOME_INST, SEQ_ATIV)
VALUES (codatuacao,
'<xsl:value-of select = "@NOME-INSTITUICAO"/>',
'<xsl:value-of select = "@SEQUENCIA-ATIVIDADE"/>'

```

```

);
<xsl:apply-templates/>
</xsl:template>

<xsl:template match = "VINCULOS»
SELECT COUNT(COD_ATUACAO)+1 INTO CODATUACAO FROM ATUA-
COES;
INSERT INTO ATUACOES(COD_ATUACAO, COD_PESSOA, TIPO)
VALUES (codatuacao,
pessoa,
'VP');
INSERT INTO VINCULO_PROF (COD_ATUACAO, SEQ_HIST, TIPO_VINCULO,
ENQ_FUNCIONAL, CARGA_HORARIA, FLAG_DEDICACAO_EXCLUSIVA,
MES_INICIO, ANO_INICIO, MES_FIM, ANO_FIM, OUTRAS_INF,
FLAG_VINCULO_EMPREGATICIO, OUTRO_VINCULO, OUTRO_
ENQUADRAMENTO)
VALUES (codatuacao,
'<xsl:value-of select = "@SEQUENCIA-HISTORICO"/>',
'<xsl:value-of select = "@TIPO-DE-VINCULO"/>',
'<xsl:value-of select = "@ENQUADRAMENTO-FUNCIONAL"/>',
'<xsl:value-of select = "@CARGA-HORARIA-SEMANAL"/>',
'<xsl:value-of select = "@FLAG-DEDICACAO-EXCLUSIVA"/>',
'<xsl:value-of select = "@MES-INICIO"/>',
'<xsl:value-of select = "@ANO-INICIO"/>',
'<xsl:value-of select = "@MES-FIM"/>',
'<xsl:value-of select = "@ANO-FIM"/>',
'<xsl:value-of select = "@OUTRAS-INFORMACOES"/>',
'<xsl:value-of select = "@FLAG-VINCULO-EMPREGATICIO"/>',
'<xsl:value-of select = "@OUTRO-VINCULO-INFORMADO"/>',
'<xsl:value-of select="@OUTRO-ENQUADRAMENTO-FUNCIONAL-
INFORMADO"/>'
);
</xsl:template>

<xsl:template match = "ATIVIDADES-DE-DIRECAO-E-ADMINISTRACAO»
<xsl:apply-templates/>
</xsl:template>

```

```

<xsl:template match = "DIRECAO-E-ADMINISTRACAO»
SELECT COUNT(COD_ATUACAO)+1 INTO CODATUACAO FROM
ATUACOES; INSERT INTO ATUACOES(COD_ATUACAO, COD_PESSOA,
TIPO)
VALUES (codatuacao,
pessoa,
'AD');
INSERT INTO ATIVIDADE_DIRECAO (COD_ATUACAO, SEQ_FUNCAO,
FLAG_PERIODO, MES_INICIO, ANO_INICIO, MES_FIM, ANO_FIM,
NOME_ORGAO, NOME_UNID, FORMATO_CARGO, CARGO_FUNCAO)
VALUES (codatuacao,
'<xsl:value-of select = "@SEQUENCIA-FUNCAO-ATIVIDADE"/>',
'<xsl:value-of select = "@FLAG-PERIODO"/>',
'<xsl:value-of select = "@MES-INICIO"/>',
'<xsl:value-of select = "@ANO-INICIO"/>',
'<xsl:value-of select = "@MES-FIM"/>',
'<xsl:value-of select = "@ANO-FIM"/>',
'<xsl:value-of select = "@NOME-ORGAO"/>',
'<xsl:value-of select = "@NOME-UNIDADE"/>',
'<xsl:value-of select = "@FORMATO-CARGO-OU-FUNCAO"/>',
'<xsl:value-of select = "@CARGO-OU-FUNCAO"/>'
);
</xsl:template>

```

```

<xsl:template match="ATIVIDADES-DE-PESQUISA-E-DESENVOLVIMENTO»
<xsl:apply-templates/>
</xsl:template>
<xsl:template match = "PESQUISA-E-DESENVOLVIMENTO»
SELECT COUNT(COD_ATUACAO)+1 INTO CODATUACAO FROM ATUA-
COES;
INSERT INTO ATUACOES(COD_ATUACAO, COD_PESSOA, TIPO)
VALUES (codatuacao,
pessoa,
'APD');
SELECT COUNT(COD_PD)+1 INTO CODPD FROM ATIVIDADE_PESQUISA;
INSERT INTO ATIVIDADE_PESQUISA (COD_PD, COD_ATUACAO,
SEQ_FUNCAO, FLAG_PERIODO, MES_INICIO, ANO_INICIO, MES_FIM,

```

```

ANO_FIM, NOME_ORGAO, NOME_UNID)
VALUES (codpd,
codatuacao,
'<xsl:value-of select = "@SEQUENCIA-FUNCAO-ATIVIDADE"/>',
'<xsl:value-of select = "@FLAG-PERIODO"/>',
'<xsl:value-of select = "@MES-INICIO"/>',
'<xsl:value-of select = "@ANO-INICIO"/>',
'<xsl:value-of select = "@MES-FIM"/>',
'<xsl:value-of select = "@ANO-FIM"/>',
'<xsl:value-of select = "@NOME-ORGAO"/>',
'<xsl:value-of select = "@NOME-UNIDADE"/>'
);
<xsl:apply-templates/>
</xsl:template>

```

```

<xsl:template match = "LINHA-DE-PESQUISA»
INSERT INTO LINHA_PESQUISA(COD_PD, SEQ_LINHA_PESQ,
TITULO_LINHA, FLAG_LINHA_ATIVA, OBJETIVOS_LINHA)
VALUES (codpd,
'<xsl:value-of select = "@SEQUENCIA-LINHA"/>',
'<xsl:value-of select = "@TITULO-DA-LINHA-DE-PESQUISA"/>',
'<xsl:value-of select = "@FLAG-LINHA-DE-PESQUISA-ATIVA"/>',
'<xsl:value-of select = "@OBJETIVOS-LINHA-DE-PESQUISA"/>'
);
</xsl:template>

```

```

<xsl:template match = "ATIVIDADES-DE-ENSINO»
<xsl:apply-templates/>
</xsl:template>
<xsl:template match = "ENSINO»
SELECT COUNT(COD_ATUACAO)+1 INTO CODATUACAO FROM
ATUACOES;INSERT INTO ATUACOES(COD_ATUACAO,COD_PESSOA,
TIPO)
VALUES (codatuacao,
pessoa,
'AE'
);

```



```

SELECT COUNT(COD_ENSINO)+1 INTO CODENSINO FROM
ATIVIDADE_ENSINO; INSERT INTO ATIVIDADE_ENSINO
(COD_ENSINO, COD_ATUACAO, SEQ_FUNCAO, FLAG_PERIODO,
TIPO_ENSINO, MES_INICIO, ANO_INICIO, MES_FIM,
ANO_FIM, NOME_ORGAO, NOME_CURSO)
VALUES(codensino,
codatuacao,
'<xsl:value-of select = "@SEQUENCIA-FUNCAO-ATIVIDADE"/>',
'<xsl:value-of select = "@FLAG-PERiodo"/>',
'<xsl:value-of select = "@TIPO-ENSINO"/>',
'<xsl:value-of select = "@MES-INICIO"/>',
'<xsl:value-of select = "@ANO-INICIO"/>',
'<xsl:value-of select = "@MES-FIM"/>',
'<xsl:value-of select = "@ANO-FIM"/>',
'<xsl:value-of select = "@NOME-ORGÃO"/>',
'<xsl:value-of select = "@NOME-CURSO"/>'
);
<xsl:apply-templates/>
</xsl:template>

<xsl:template match = "DISCIPLINA»
INSERT INTO DISCIPLINA (COD_ENSINO, SEQ_ESPECIFICACAO,
COMENTARIO)
VALUES (codensino,
'<xsl:value-of select = "@SEQUENCIA-ESPECIFICACAO"/>',
'<xsl:value-of select = "."/>'
);
</xsl:template>

<xsl:template match = "ATIVIDADES-DE-SERVICO-TECNICO-
ESPECIALIZADO»
<xsl:apply-templates/>
</xsl:template>
<xsl:template match = "SERVICO-TECNICO-ESPECIALIZADO»
SELECT COUNT(COD_ATUACAO)+1 INTO CODATUACAO FROM ATUA-
COES;
INSERT INTO ATUACOES(COD_ATUACAO, COD_PESSOA, TIPO)

```

```

VALUES (codatuacao,
        pessoa,
        'ST');
INSERT INTO SERVICOS_TECNICOS (COD_ATUACAO, SEQ_FUNCAO,
FLAG_PERIODO, MES_INICIO, ANO_INICIO, MES_FIM, ANO_FIM,
NOME_ORGAO, NOME_UNID, SERVICO_REALIZADO)
VALUES (codatuacao,
        '<xsl:value-of select = "@SEQUENCIA-FUNCAO-ATIVIDADE"/>',
        '<xsl:value-of select = "@FLAG-PERIODO"/>',
        '<xsl:value-of select = "@MES-INICIO"/>',
        '<xsl:value-of select = "@ANO-INICIO"/>',
        '<xsl:value-of select = "@MES-FIM"/>',
        '<xsl:value-of select = "@ANO-FIM"/>',
        '<xsl:value-of select = "@NOME-ORGAO"/>',
        '<xsl:value-of select = "@NOME-UNIDADE"/>',
        '<xsl:value-of select = "@SERVICO-REALIZADO"/>'
);
</xsl:template>

<xsl:template match="ATIVIDADES-DE-EXTENSAO-UNIVERSITARIA»
<xsl:apply-templates/>
</xsl:template>
<xsl:template match = "EXTENSAO-UNIVERSITARIA»
SELECT COUNT(COD_ATUACAO)+1 INTO CODATUACAO FROM ATUA-
COES;
INSERT INTO ATUACOES(COD_ATUACAO, COD_PESSOA, TIPO)
VALUES (codatuacao,
        pessoa,
        'EU');
INSERT INTO ATIVIDADES_EXTENSAO (COD_ATUACAO, SEQ_FUNCAO,
FLAG_PERIODO, MES_INICIO, ANO_INICIO, MES_FIM, ANO_FIM,
NOME_ORGAO, NOME_UNID, ATIVIDADE_EXTENSAO)
VALUES (codatuacao,
        '<xsl:value-of select = "@SEQUENCIA-FUNCAO-ATIVIDADE"/>',
        '<xsl:value-of select = "@FLAG-PERIODO"/>',
        '<xsl:value-of select = "@MES-INICIO"/>',
        '<xsl:value-of select = "@ANO-INICIO"/>',

```

```

'<xsl:value-of select = "@MES-FIM"/>',
'<xsl:value-of select = "@ANO-FIM"/>',
'<xsl:value-of select = "@NOME-ORGAO"/>',
'<xsl:value-of select = "@NOME-UNIDADE"/>',
'<xsl:value-of select="@ATIVIDADE-DE-EXTENSAO-REALIZADA"/>'
);
</xsl:template>
<xsl:template match="ATIVIDADES-DE-TREINAMENTO-MINISTRADO»
<xsl:apply-templates/>
</xsl:template>

<xsl:template match = "TREINAMENTO-MINISTRADO»
SELECT COUNT(COD_ATUACAO)+1 INTO CODATUACAO FROM
ATUACOES;INSERT INTO ATUACOES(COD_ATUACAO, COD_PESSOA,
TIPO)
VALUES (codatuacao,
pessoa,
'TM');
SELECT COUNT(COD_TREINAMENTO)+1 INTO CODTREINAMENTO FROM
TREINAMENTOS_MINISTRADOS;
INSERT INTO TREINAMENTOS_MINISTRADOS (COD_TREINAMENTO,
COD_ATUACAO, SEQ_FUNCAO, FLAG_PERIODO, MES_INICIO,
ANO_INICIO, MES_FIM, ANO_FIM, NOME_ORGAO, NOME_UNID)
VALUES (codtreinamento,
codatuacao,
'<xsl:value-of select="@SEQUENCIA-FUNCAO-ATIVIDADE"/>',
'<xsl:value-of select = "@FLAG-PERiodo"/>',
'<xsl:value-of select = "@MES-INICIO"/>',
'<xsl:value-of select = "@ANO-INICIO"/>',
'<xsl:value-of select = "@MES-FIM"/>',
'<xsl:value-of select = "@ANO-FIM"/>',
'<xsl:value-of select = "@NOME-ORGAO"/>',
'<xsl:value-of select = "@NOME-UNIDADE"/>'
);
<xsl:apply-templates/>
</xsl:template>

```

```

<xsl:template match = "TREINAMENTO»
INSERT INTO TREINAMENTO(COD_TREINAMENTO,SEQ_ESPECIFICACAO,
OBSERVACAO)
VALUES (codtreinamento,
'<xsl:value-of select = "@SEQUENCIA-ESPECIFICACAO"/>',
'<xsl:value-of select = "."/>'
);
</xsl:template>

```

```

<xsl:template match = "OUTRAS-ATIVIDADES-TECNICO-CIENTIFICA»
<xsl:apply-templates/>
</xsl:template>
<xsl:template match = "OUTRA-ATIVIDADE-TECNICO-CIENTIFICA»
SELECT COUNT(COD_ATUACAO)+1 INTO CODATUACAO FROM ATUA-
COES;
INSERT INTO ATUACOES(COD_ATUACAO, COD_PESSOA, TIPO)
VALUES (codatuacao,
pessoa,
'OUTRAS');
INSERT INTO OUTRAS_ATIVIDADES (COD_ATUACAO, SEQ_FUNCAO,
FLAG_PERIODO, MES_INICIO, ANO_INICIO, MES_FIM, ANO_FIM,
NOME_ORGAO, NOME_UNID, ATIVIDADE_REALIZADA)
VALUES (codatuacao,
'<xsl:value-of select = "@SEQUENCIA-FUNCAO-ATIVIDADE"/>',
'<xsl:value-of select = "@FLAG-PERIODO"/>',
'<xsl:value-of select = "@MES-INICIO"/>',
'<xsl:value-of select = "@ANO-INICIO"/>',
'<xsl:value-of select = "@MES-FIM"/>',
'<xsl:value-of select = "@ANO-FIM"/>',
'<xsl:value-of select = "@NOME-ORGAO"/>',
'<xsl:value-of select = "@NOME-UNIDADE"/>',
'<xsl:value-of select = "@ATIVIDADE-REALIZADA"/>'
);
</xsl:template>

```

```

<xsl:template match = "PRODUCAO-BIBLIOGRAFICA»
<xsl:apply-templates>

```

```

</xsl:template>

<xsl:template match = "TRABALHOS-EM-EVENTOS»
<xsl:apply-templates/>
</xsl:template>
<xsl:template match = "TRABALHO-EM-EVENTOS»
SELECT COUNT(COD_TRAB)+1 INTO CODTRAB FROM TRABALHOS_
EM_EVENTOS;INSERT INTO TRABALHOS_EM_EVENTOS(COD_TRAB,
COD_PESSOA, SEQ_PRODUCAO, NATUREZA, TITULO, ANO,
PAIS_EVENTO, IDIOMA, MEIO_DIVULGACAO, HOME_PAGE,
FLAG_RELEVANCIA)
VALUES (codtrab,
pessoa,
'<xsl:value-of select="@SEQUENCIA-PRODUCAO"/>',
'<xsl:value-of select="DADOS-BASICOS-DO-TRABALHO/@NATUREZA"/>',
'<xsl:value-of select="DADOS-BASICOS-DO-TRABALHO/@TITULO-DO-
TRABALHO"/>',
'<xsl:value-of select="DADOS-BASICOS-DO-TRABALHO/@ANO-DO-
TRABALHO"/>',
'<xsl:value-of select="DADOS-BASICOS-DO-TRABALHO/@PAIS-DO-
EVENTO"/>',
'<xsl:value-of select="DADOS-BASICOS-DO-TRABALHO/@IDIOMA"/>',
'<xsl:value-of select="DADOS-BASICOS-DO-TRABALHO/@MEIO-DE-
DIVULGACAO"/>',
'<xsl:value-of select="DADOS-BASICOS-DO-TRABALHO/@HOME-PAGE-DO-
TRABALHO"/>',
'<xsl:value-of select="DADOS-BASICOS-DO-TRABALHO/@FLAG-
RELEVANCIA"/>'
);
INSERT INTO DETALHAMENTO_TRAB (COD_TRAB, CLASSIFICACAO_
EVENTO,NOME_EVENTO, CIDADE_EVENTO, ANO_REALIZACAO,
TITULO_ANAIS, VOLUME, FASCICULO, SERIE, PAGINA_INICIAL,
PAGINA_FINAL, ISBN, EDITORA, CIDADE_EDITORA)
VALUES (codtrab,
'<xsl:value-of select="DETALHAMENTO-DO-TRABALHO/@CLASSIFICACAO-
DO-EVENTO"/>',
'<xsl:value-of select="DETALHAMENTO-DO-TRABALHO/@NOME-DO-EVENTO"/>',

```

```
'<xsl:value-of select="DETALHAMENTO-DO-TRABALHO/@CIDADE-DO-
EVENTO"/>',
'<xsl:value-of select="DETALHAMENTO-DO-TRABALHO/@ANO-DE-
REALIZACAO"/>',
'<xsl:value-of select="DETALHAMENTO-DO-TRABALHO/@TITULO-DOS-ANAIS-
OU-PROCEEDINGS"/>',
'<xsl:value-of select="DETALHAMENTO-DO-TRABALHO/@VOLUME"/>',
'<xsl:value-of select="DETALHAMENTO-DO-TRABALHO/@FASCICULO"/>',
'<xsl:value-of select="DETALHAMENTO-DO-TRABALHO/@SERIE"/>',
'<xsl:value-of select="DETALHAMENTO-DO-TRABALHO/@PAGINA-INICIAL"/>',
'<xsl:value-of select="DETALHAMENTO-DO-TRABALHO/@PAGINA-FINAL"/>',
'<xsl:value-of select="DETALHAMENTO-DO-TRABALHO/@ISBN"/>',
'<xsl:value-of select="DETALHAMENTO-DO-TRABALHO/@NOME-DA-
EDITORIA"/>',
'<xsl:value-of select="DETALHAMENTO-DO-TRABALHO/@CIDADE-DA-
EDITORIA"/>'
);
```

```
<xsl:for-each select = "AUTORES">
INSERT INTO AUTORES_TRAB (COD_TRAB, NOME_AUTOR, NOME_
CITACAO, ORDEM_AUTORIA)
VALUES (codtrab,
'<xsl:value-of select = "@NOME-COMPLETO-DO-AUTOR"/>',
'<xsl:value-of select = "@NOME-PARA-CITACAO"/>',
'<xsl:value-of select = "@ORDEM-DE-AUTORIA"/>'
);
</xsl:for-each>
INSERT INTO PALAVRAS_CHAVE_TRAB (COD_TRAB, PALAVRA1,
PALAVRA2, PALAVRA3, PALAVRA4, PALAVRA5, PALAVRA6)
VALUES (codtrab,
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-1"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-2"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-3"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-4"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-5"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-6"/>'
);
```

```

<xsl:for-each select = "AREAS-DO-CONHECIMENTO/*»
SELECT COUNT(COD_AREA)+1 INTO CODAREA FROM AREA_DO_
CONHECIMENTO_TRAB; INSERT INTO AREA_DO_CONHECIMENTO_
TRAB(COD_TRAB, COD_AREA, GRANDE_AREA, AREA, SUB_AREA,
ESPECIALIDADE)
VALUES (codtrab,
codarea,
'<xsl:value-of select = "@NOME-GRANDE-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select = "@NOME-DA-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select = "@NOME-DA-SUB-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select = "@NOME-DA-ESPECIALIDADE"/>'
);
</xsl:for-each>
</xsl:template>

```

```

<xsl:template match = "ARTIGOS-PUBLICADOS»
<xsl:apply-templates/>
</xsl:template>
<xsl:template match = "ARTIGO-PUBLICADO»
SELECT COUNT(COD_ARTIGO)+1 INTO CODARTIGO FROM ARTIGOS_
PUBLICADOS; INSERT INTO ARTIGOS_PUBLICADOS (COD_ARTIGO,
COD_PESSOA, SEQ_PRODUCAO, NATUREZA, TITULO, ANO, PAIS_
PUBLICACAO,IDIOMA, MEIO_DIVULGACAO, HOME_PAGE,
FLAG_RELEVANCIA)
VALUES (codartigo,
pessoa,
'<xsl:value-of select = "@SEQUENCIA-PRODUCAO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DO-ARTIGO/@NATUREZA"/>',
'<xsl:value-of select = "DADOS-BASICOS-DO-ARTIGO/@TITULO-DO-
ARTIGO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DO-ARTIGO/@ANO-DO-
ARTIGO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DO-ARTIGO/@PAIS-
DE-PUBLICACAO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DO-ARTIGO/@IDIOMA"/>',
'<xsl:value-of select = "DADOS-BASICOS-DO-ARTIGO/@MEIO-

```

```

DE-DIVULGACAO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DO-ARTIGO/@HOME-
PAGE-DO-TRABALHO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DO-ARTIGO/@FLAG-
RELEVANCIA"/>'
);
INSERT INTO DETALHAMENTO_ARTIGO (COD_ARTIGO, TITULO_PERIODICO_
REVISTA, ISSN, VOLUME, FASCICULO, SERIE, PAGINA_INICIAL,
PAGINA_FINAL, LOCAL_PUBLICACAO)
VALUES (codartigo,
'<xsl:value-of select="DETALHAMENTO-DO-ARTIGO/@TITULO-DO-
PERIODICO-OU-REVISTA"/>',
'<xsl:value-of select="DETALHAMENTO-DO-ARTIGO/@ISSN"/>',
'<xsl:value-of select="DETALHAMENTO-DO-ARTIGO/@VOLUME"/>',
'<xsl:value-of select="DETALHAMENTO-DO-ARTIGO/@FASCICULO"/>',
'<xsl:value-of select="DETALHAMENTO-DO-ARTIGO/@SERIE"/>',
'<xsl:value-of select="DETALHAMENTO-DO-ARTIGO/@PAGINA-INICIAL"/>',
'<xsl:value-of select="DETALHAMENTO-DO-ARTIGO/@PAGINA-FINAL"/>',
'<xsl:value-of select="DETALHAMENTO-DO-ARTIGO/@LOCAL-DE-
PUBLICACAO"/>'
);
<xsl:for-each select = "AUTORES»
INSERT INTO AUTORES_ARTIGO (COD_ARTIGO, NOME, NOME_CITACAO,
ORDEM_AUTORIA)
VALUES (codartigo,
'<xsl:value-of select = "@NOME-COMPLETO-DO-AUTOR"/>',
'<xsl:value-of select = "@NOME-PARA-CITACAO"/>',
'<xsl:value-of select = "@ORDEM-DE-AUTORIA"/>'
);
</xsl:for-each>
INSERT INTO PALAVRAS_CHAVE_ARTIGO (COD_ARTIGO, PALAVRA1,
PALAVRA2, PALAVRA3, PALAVRA4, PALAVRA5, PALAVRA6)
VALUES (codartigo,
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-1"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-2"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-3"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-4"/>',

```



```

'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-5"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-6"/>'
);
<xsl:for-each select = "AREAS-DO-CONHECIMENTO/*»
SELECT COUNT(COD_AREA)+1 INTO CODAREA FROM AREA_
CONHECIMENTO_ARTIGO; INSERT INTO AREA_CONHECIMENTO_
ARTIGO (COD_ARTIGO, COD_AREA, GRANDE_AREA, AREA,
SUB_AREA, ESPECIALIDADE)
VALUES (codartigo,
codarea,
'<xsl:value-of select = "@NOME-GRANDE-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select = "@NOME-DA-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select = "@NOME-DA-SUB-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select = "@NOME-DA-ESPECIALIDADE"/>'
);
</xsl:for-each>
INSERT INTO SETOR_ATIVIDADE_ARTIGO (COD_ARTIGO, SETOR1,
SETOR2,SETOR3)
VALUES (codartigo,
'<xsl:value-of select = "SETORES-DE-ATIVIDADE/@SETOR-DE-
ATIVIDADE-1"/>',
'<xsl:value-of select = "SETORES-DE-ATIVIDADE/@SETOR-DE-
ATIVIDADE-2"/>',
'<xsl:value-of select = "SETORES-DE-ATIVIDADE/@SETOR-DE-
ATIVIDADE-3"/>'
);
INSERT INTO INFORMACOES_ADICIONAIS_ARTIGO (COD_ARTIGO,
INFORMACOES)
VALUES (codartigo,
'<xsl:value-of select = "INFORMACOES-ADICIONAIS/@DESCRICAO-
INFORMACOES-ADICIONAIS"/>'
);
</xsl:template>

<xsl:template match="DEMAIS-TIPOS-DE-PRODUCAO-BIBLIOGRAFICA»
<xsl:apply-templates/>
</xsl:template>

```

```

<xsl:template match = "OUTRA-PRODUCAO-BIBLIOGRAFICA»
SELECT COUNT(COD_PRODUCAO)+1 INTO CODPRODUCAO FROM
OUTRA_PROD_BIBLI; INSERT INTO OUTRA_PROD_BIBLI
(COD_PRODUCAO, COD_PESSOA, SEQ_PRODUCAO, NATUREZA, TITULO,
ANO, PAIS_PUBLICACAO, IDIOMA, MEIO_DIVULGACAO, HOME_PAGE,
FLAG_RELEVANCIA)
VALUES (codproducao,
pessoa,
'<xsl:value-of select = "@SEQUENCIA-PRODUCAO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-OUTRA-PRODUCAO/
@NATUREZA"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-OUTRA-PRODUCAO/@
TITULO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-OUTRA-PRODUCAO/@ANO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-OUTRA-PRODUCAO/@PAIS-
DE
-PUBLICACAO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-OUTRA-PRODUCAO/@IDIOMA"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-OUTRA-PRODUCAO/@MEIO-
DE-
DIVULGACAO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-OUTRA-PRODUCAO/@HOME-
PAGE-DO-TRABALHO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-OUTRA-PRODUCAO/@FLAG-
RELEVANCIA"/>'
);
INSERT INTO DETALHAMENTO_OUTRA_PROD (COD_PRODUCAO, EDI-
TORA,
CIDADE_EDITORA, NUMERO_PAGINAS, ISSN_ISBN)
VALUES (codproducao,
'<xsl:value-of select = "DETALHAMENTO-DE-OUTRA-PRODUCAO/@EDITORA"/>',
'<xsl:value-of select = "DETALHAMENTO-DE-OUTRA-PRODUCAO/@CIDADE-
DA-EDITORA"/>',
'<xsl:value-of select = "DETALHAMENTO-DE-OUTRA-PRODUCAO/@NUMERO-

```

```

DE-PAGINAS"/>',
'<xsl:value-of select = "DETALHAMENTO-DE-OUTRA-PRODUCAO/@ISSN-
ISBN"/>'
);
INSERT INTO AUTORES_OUTRA_PROD (COD_PRODUCAO, NOME_AUTOR,
NOME_CITACAO, ORDEM_AUTORIA)
VALUES (codproducao,
'<xsl:value-of select = "AUTORES/@NOME-COMPLETO-DO-AUTOR"/>',
'<xsl:value-of select = "AUTORES/@NOME-PARA-CITACAO"/>',
'<xsl:value-of select = "AUTORES/@ORDEM-DE-AUTORIA"/>'
);
INSERT INTO PALAVRAS_CHAVE_OUTRA_PROD (COD_PRODUCAO, PA-
LAVRA1, PALAVRA2,
PALAVRA3, PALAVRA4, PALAVRA5, PALAVRA6)
VALUES (codproducao,
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-1"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-2"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-3"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-4"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-5"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-6"/>'
);
<xsl:for-each select = "AREAS-DO-CONHECIMENTO/*">
SELECT COUNT(COD_AREA)+1 INTO CODAREA FROM AREA_
CONHECIMENTO_OUTRA_PROD; INSERT INTO AREA_CONHECIMENTO_
OUTRA_PROD (COD_PRODUCAO, COD_AREA,GRANDE_AREA,
AREA, SUB_AREA, ESPECIALIDADE)
VALUES (codproducao,
codarea,
'<xsl:value-of select = "@NOME-GRANDE-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select = "@NOME-DA-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select = "@NOME-DA-SUB-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select = "@NOME-DA-ESPECIALIDADE"/>'
);
</xsl:for-each>
INSERT INTO SETOR_ATIV_OUTRA_PROD (COD_PRODUCAO, SETOR1,
SETOR2, SETOR3)

```

```

VALUES (codproducao,
'<xsl:value-of select = "SETORES-DE-ATIVIDADE/@SETOR-DE-
ATIVIDADE-1"/>',
'<xsl:value-of select = "SETORES-DE-ATIVIDADE/@SETOR-DE-
ATIVIDADE-2"/>',
'<xsl:value-of select = "SETORES-DE-ATIVIDADE/@SETOR-DE-
ATIVIDADE-3"/>'
);
INSERT INTO INF_ADICIONAIS_OUTRA_PROD (COD_PRODUCAO,
INFORMACOES)
VALUES (codproducao,
'<xsl:value-of select="INFORMACOES-ADICIONAIS/@DESCRICAO-
INFORMACOES-ADICIONAIS"/>'
);
</xsl:template>

<xsl:template match = "PRODUCAO-TECNICA»
<xsl:apply-templates/>
</xsl:template>
<xsl:template match = "TRABALHO-TECNICO»
SELECT COUNT(COD_TRAB_TEC)+1 INTO CODTRABTEC FROM
TRABALHOS_TECNICOS; INSERT INTO TRABALHOS_TECNICOS
(COD_TRAB_TEC, COD_PESSOA, SEQ_PRODUCAO, NATUREZA,
TITULO, ANO, PAIS, IDIOMA, MEIO_DIVULGACAO, HOME_PAGE,
FLAG_RELEVANCIA)
VALUES (codtrabtec,
pessoa,
'<xsl:value-of select = "@SEQUENCIA-PRODUCAO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DO-TRABALHO-
TECNICO/@NATUREZA"/>',
'<xsl:value-of select = "DADOS-BASICOS-DO-TRABALHO-
TECNICO/@TITULO-DO-TRABALHO-TECNICO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DO-TRABALHO-
TECNICO/@ANO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DO-TRABALHO-
TECNICO/@PAIS"/>',
'<xsl:value-of select = "DADOS-BASICOS-DO-TRABALHO-

```

```

TECNICO/@IDIOMA"/>',
'<xsl:value-of select = "DADOS-BASICOS-DO-TRABALHO-
TECNICO/@MEIO-DE-DIVULGACAO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DO-TRABALHO-
TECNICO/@HOME-PAGE-DO-TRABALHO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DO-TRABALHO-
TECNICO/@FLAG-RELEVANCIA"/>'
);
INSERT INTO DETALHAMENTO_TRAB_TEC(COD_TRAB_TEC,FINALIDADE,
DURACAO_MESES, NUMERO_PAGINAS, DISPONIBILIDADE,
INST_FINANCIADORA, CIDADE)
VALUES (codtrabtec,
'<xsl:value-of select="DETALHAMENTO-DO-TRABALHO-TECNICO/
@FINALIDADE"/>',
'<xsl:value-of select="DETALHAMENTO-DO-TRABALHO-TECNICO/
@DURACAO-EM-MESES"/>',
'<xsl:value-of select="DETALHAMENTO-DO-TRABALHO-TECNICO/
@NUMERO-DE-PAGINAS"/>',
'<xsl:value-of select="DETALHAMENTO-DO-TRABALHO-TECNICO/
@DISPONIBILIDADE"/>',
'<xsl:value-of select="DETALHAMENTO-DO-TRABALHO-TECNICO/
@INSTITUICAO-FINANCIADORA"/>',
'<xsl:value-of select="DETALHAMENTO-DO-TRABALHO-TECNICO/
@CIDADE-DO-TRABALHO"/>'
);
<xsl:for-each select = "AUTORES»
INSERT INTO AUTORES_TRAB_TEC(COD_TRAB_TEC, NOME_AUTOR,
NOME_CITACAO, ORDEM_AUTORIA)
VALUES (codtrabtec,
'<xsl:value-of select = "@NOME-COMPLETO-DO-AUTOR"/>',
'<xsl:value-of select = "@NOME-PARA-CITACAO"/>',
'<xsl:value-of select = "@ORDEM-DE-AUTORIA"/>'
);
</xsl:for-each>
INSERT INTO PALAVRAS_CHAVE_TRAB_TEC (COD_TRAB_TEC,
PALAVRA1, PALAVRA2, PALAVRA3, PALAVRA4, PALAVRA5, PALAVRA6)
VALUES (codtrabtec,

```

```

'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-1"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-2"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-3"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-4"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-5"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-6"/>'
);
<xsl:for-each select = "AREAS-DO-CONHECIMENTO/*">
SELECT COUNT(COD_AREA)+1 INTO CODAREA FROM AREA_
CONHECIMENTO_TRAB_TEC;
INSERT INTO AREA_CONHECIMENTO_TRAB_TEC (COD_TRAB_TEC,
COD_AREA, GRANDE_AREA, AREA, SUB_AREA, ESPECIALIDADE)
VALUES (codtrabtec,
codarea,
'<xsl:value-of select = "@NOME-GRANDE-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select = "@NOME-DA-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select = "@NOME-DA-SUB-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select = "@NOME-DA-ESPECIALIDADE"/>'
);
</xsl:for-each>
INSERT INTO SETOR_ATIVIDADE_TRAB_TEC (COD_TRAB_TEC, SETOR1,
SETOR2, SETOR3)
VALUES (codtrabtec,
'<xsl:value-of select = "SETORES-DE-ATIVIDADE/@SETOR-DE-ATIVIDADE-
1"/>',
'<xsl:value-of select = "SETORES-DE-ATIVIDADE/@SETOR-DE-ATIVIDADE-
2"/>',
'<xsl:value-of select = "SETORES-DE-ATIVIDADE/@SETOR-DE-ATIVIDADE-
3"/>'
);
INSERT INTO INF_ADICIONAIS_TRAB_TEC (COD_TRAB_TEC, INFORMA-
COES)
VALUES (codtrabtec,
'<xsl:value-of select="INFORMACOES-ADICIONAIS/@DESCRICA-
O-INFORMACOES-ADICIONAIS"/>'
);
</xsl:template>

```

```

<xsl:template match = "OUTRA-PRODUCAO»
<xsl:apply-templates/>
</xsl:template>
<xsl:template match = "ORIENTACOES-CONCLUIDAS»
<xsl:apply-templates/>
</xsl:template>
<xsl:template match = "OUTRAS-ORIENTACOES-CONCLUIDAS»
SELECT COUNT(COD_ORIENTACAO)+1 INTO CODORIENTACAO FROM
ORIENTACOES_CONCLUIDAS; INSERT INTO ORIENTACOES_
CONCLUIDAS (COD_ORIENTACAO, COD_PESSOA, SEQ_PRODUCAO,
NATUREZA, TIPO, TITULO, ANO, PAIS, IDIOMA, HOME_PAGE,
FLAG_RELEVANCIA)
VALUES (codorientacao,
pessoa,
'<xsl:value-of select = "@SEQUENCIA-PRODUCAO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-OUTRAS-
ORIENTACOES-CONCLUIDAS/@NATUREZA"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-OUTRAS-
ORIENTACOES-CONCLUIDAS/@TIPO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-OUTRAS-
ORIENTACOES-CONCLUIDAS/@TITULO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-OUTRAS-
ORIENTACOES-CONCLUIDAS/@ANO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-OUTRAS-
ORIENTACOES-CONCLUIDAS/@PAIS"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-OUTRAS-
ORIENTACOES-CONCLUIDAS/@IDIOMA"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-OUTRAS-
ORIENTACOES-CONCLUIDAS/@HOME-PAGE"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-OUTRAS-
ORIENTACOES-CONCLUIDAS/@FLAG-RELEVANCIA"/>'
);
INSERT INTO DETALHAMENTO_ORIENTACOES (COD_ORIENTACAO,
NOME_ORIENTADO, NOME_INST, NOME_CURSO, FLAG_BOLSA,
NOME_AGENCIA, TIPO_ORIENTACAO, NUMERO_PAGINAS)
VALUES (codorientacao,

```

```

'<xsl:value-of select = "DETALHAMENTO-DE-OUTRAS-
ORIENTACOES-CONCLUIDAS/@NOME-DO-ORIENTADO"/>',
'<xsl:value-of select = "DETALHAMENTO-DE-OUTRAS-
ORIENTACOES-CONCLUIDAS/@NOME-DA-INSTITUICAO"/>',
'<xsl:value-of select = "DETALHAMENTO-DE-OUTRAS-
ORIENTACOES-CONCLUIDAS/@NOME-DO-CURSO"/>',
'<xsl:value-of select = "DETALHAMENTO-DE-OUTRAS-
ORIENTACOES-CONCLUIDAS/@FLAG-BOLSA"/>',
'<xsl:value-of select = "DETALHAMENTO-DE-OUTRAS-
ORIENTACOES-CONCLUIDAS/@NOME-DA-AGENCIA"/>',
'<xsl:value-of select = "DETALHAMENTO-DE-OUTRAS-
ORIENTACOES-CONCLUIDAS/@TIPO-DE-ORIENTACAO-CONCLUIDA"/>',
'<xsl:value-of select = "DETALHAMENTO-DE-OUTRAS-
ORIENTACOES-CONCLUIDAS/@NUMERO-DE-PAGINAS"/>'
);
INSERT INTO PALAVRAS_CHAVE_ORIENTACOES (COD_ORIENTACAO,
PALAVRA1, PALAVRA2, PALAVRA3, PALAVRA4, PALAVRA5, PALAVRA6)
VALUES (codorientacao,
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-1"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-2"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-3"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-4"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-5"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-6"/>'
);
<xsl:for-each select = "AREAS-DO-CONHECIMENTO/*">
SELECT COUNT(COD_AREA)+1 INTO CODAREA FROM AREA_
CONHECIMENTO_ORIENTACOES; INSERT INTO AREA_CONHECIMENTO_
ORIENTACOES (COD_ORIENTACAO, COD_AREA, GRANDE_AREA, AREA,
SUB_AREA, ESPECIALIDADE)
VALUES (codorientacao,
codarea,
'<xsl:value-of select = "@NOME-GRANDE-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select = "@NOME-DA-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select = "@NOME-DA-SUB-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select = "@NOME-DA-ESPECIALIDADE"/>'
);

```



```

</xsl:for-each>
INSERT INTO SETOR_ATIV_ORIENTACOES (COD_ORIENTACAO, SETOR1,
SETOR2, SETOR3)
VALUES (codorientacao,
'<xsl:value-of select = "SETORES-DE-ATIVIDADE/@SETOR-DE-ATIVIDADE-
1"/>',
'<xsl:value-of select = "SETORES-DE-ATIVIDADE/@SETOR-DE-ATIVIDADE-
2"/>',
'<xsl:value-of select = "SETORES-DE-ATIVIDADE/@SETOR-DE-ATIVIDADE-
3"/>'
);
INSERT INTO INF_ADICIONAIS_ORIENTACOES (COD_ORIENTACAO,
INFORMACOES)
VALUES (codorientacao,
'<xsl:value-of select = "@INFORMACOES-ADICIONAIS/
DESCRICAO-INFORMACOES-ADICIONAIS"/>'
);
</xsl:template>

```

```

<xsl:template match = "ORIENTACOES-CONCLUIDAS-PARA-MESTRADO»
SELECT COUNT(COD_ORIENTACAO)+1 INTO CODORIENTACAO FROM
ORIENTACOES_CONCLUIDAS; INSERT INTO ORIENTACOES_
CONCLUIDAS (COD_ORIENTACAO, COD_PESSOA, SEQ_PRODUCAO,
NATUREZA, TIPO, TITULO, ANO, PAIS, IDIOMA, HOME_PAGE,
FLAG_RELEVANCIA)
VALUES (codorientacao,
pessoa,
'<xsl:value-of select = "@SEQUENCIA-PRODUCAO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-ORIENTACOES-
CONCLUIDAS-PARA-MESTRADO/@NATUREZA"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-ORIENTACOES-
CONCLUIDAS-PARA-MESTRADO/@TIPO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-ORIENTACOES-
CONCLUIDAS-PARA-MESTRADO/@TITULO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-ORIENTACOES-
CONCLUIDAS-PARA-MESTRADO/@ANO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-ORIENTACOES-

```

```

CONCLUIDAS-PARA-MESTRADO/@PAIS"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-ORIENTACOES-
CONCLUIDAS-PARA-MESTRADO/@IDIOMA"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-ORIENTACOES-
CONCLUIDAS-PARA-MESTRADO/@HOME-PAGE"/>',
'<xsl:value-of select = "DADOS-BASICOS-DE-ORIENTACOES-
CONCLUIDAS-PARA-MESTRADO/@FLAG-RELEVANCIA"/>'
);
INSERT INTO DETALHAMENTO_ORIENTACOES (COD_ORIENTACAO,
NOME_ORIENTADO,
NOME_INST, NOME_CURSO, FLAG_BOLSA, NOME_AGENCIA, TIPO_
ORIENTACAO, NUMERO_PAGINAS)
VALUES (codorientacao,
'<xsl:value-of select = "DETALHAMENTO-DE-ORIENTACOES-
CONCLUIDAS-PARA-MESTRADO/@NOME-DO-ORIENTADO"/>',
'<xsl:value-of select = "DETALHAMENTO-DE-ORIENTACOES-
CONCLUIDAS-PARA-MESTRADO/@NOME-DA-INSTITUICAO"/>',
'<xsl:value-of select = "DETALHAMENTO-DE-ORIENTACOES-
CONCLUIDAS-PARA-MESTRADO/@NOME-DO-CURSO"/>',
'<xsl:value-of select = "DETALHAMENTO-DE-ORIENTACOES-
CONCLUIDAS-PARA-MESTRADO/@FLAG-BOLSA"/>',
'<xsl:value-of select = "DETALHAMENTO-DE-ORIENTACOES-
CONCLUIDAS-PARA-MESTRADO/@NOME-DA-AGENCIA"/>',
'<xsl:value-of select = "DETALHAMENTO-DE-ORIENTACOES-
CONCLUIDAS-PARA-MESTRADO/@TIPO-DE-ORIENTACAO-CONCLUIDA"/>',
'<xsl:value-of select = "DETALHAMENTO-DE-ORIENTACOES-
CONCLUIDAS-PARA-MESTRADO/@NUMERO-DE-PAGINAS"/>'
);
INSERT INTO PALAVRAS_CHAVE_ORIENTACOES (COD_ORIENTACAO,
PALAVRA1, PALAVRA2, PALAVRA3, PALAVRA4, PALAVRA5, PALAVRA6)
VALUES (codorientacao,
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-1"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-2"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-3"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-4"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-5"/>',
'<xsl:value-of select = "PALAVRAS-CHAVE/@PALAVRA-CHAVE-6"/>'

```

```

);
<xsl:for-each select = "AREAS-DO-CONHECIMENTO/*">
SELECT COUNT(COD_AREA)+1 INTO CODAREA FROM AREA_
CONHECIMENTO_ORIENTACOES; INSERT INTO AREA_CONHECIMENTO_
ORIENTACOES (COD_ORIENTACAO, COD_AREA, GRANDE_AREA, AREA,
SUB_AREA, ESPECIALIDADE)
VALUES (codorientacao,
codarea,
'<xsl:value-of select = "@NOME-GRANDE-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select = "@NOME-DA-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select = "@NOME-DA-SUB-AREA-DO-CONHECIMENTO"/>',
'<xsl:value-of select = "@NOME-DA-ESPECIALIDADE"/>'
);
</xsl:for-each>
INSERT INTO SETOR_ATIV_ORIENTACOES (COD_ORIENTACAO, SETOR1,
SETOR2, SETOR3)
VALUES (codorientacao,
'<xsl:value-of select = "SETORES-DE-ATIVIDADE/@SETOR-DE-ATIVIDADE-
1"/>',
'<xsl:value-of select = "SETORES-DE-ATIVIDADE/@SETOR-DE-ATIVIDADE-
2"/>',
'<xsl:value-of select = "SETORES-DE-ATIVIDADE/@SETOR-DE-ATIVIDADE-
3"/>'
);
INSERT INTO INF_ADICIONAIS_ORIENTACOES (COD_ORIENTACAO,
INFORMACOES)
VALUES (codorientacao,
'<xsl:value-of select = "@INFORMACOES-ADICIONAIS/
DESCRICAO-INFORMACOES-ADICIONAIS"/>'
);
</xsl:template>

<xsl:template match = "DADOS-COMPLEMENTARES">
<xsl:apply-templates/>
</xsl:template>
<xsl:template match = "PARTICIPACAO-EM-BANCA-TRABALHOS-
CONCLUSAO">

```

```

<xsl:apply-templates/>
</xsl:template>
<xsl:template match = "PARTICIPACAO-EM-BANCA-DE-MESTRADO»
SELECT COUNT(COD_PARTICIPACAO)+1 INTO CODPARTICIPACAO FROM
PARTICIPACAO_BANCA;
INSERT INTO PARTICIPACAO_BANCA (COD_PARTICIPACAO,
COD_PESSOA, SEQ_PRODUCAO, NATUREZA, TIPO, TITULO, ANO,
PAIS, IDIOMA, HOME_PAGE)
VALUES (codparticipacao,
pessoa,
'<xsl:value-of select = "@SEQUENCIA-PRODUCAO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DA-PARTICIPACAO-
EM-BANCA-DE-MESTRADO/@NATUREZA"/>',
'<xsl:value-of select = "DADOS-BASICOS-DA-PARTICIPACAO-
EM-BANCA-DE-MESTRADO/@TIPO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DA-PARTICIPACAO-
EM-BANCA-DE-MESTRADO/@TITULO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DA-PARTICIPACAO-
EM-BANCA-DE-MESTRADO/@ANO"/>',
'<xsl:value-of select = "DADOS-BASICOS-DA-PARTICIPACAO-
EM-BANCA-DE-MESTRADO/@PAIS"/>',
'<xsl:value-of select = "DADOS-BASICOS-DA-PARTICIPACAO-
EM-BANCA-DE-MESTRADO/@IDIOMA"/>',
'<xsl:value-of select = "DADOS-BASICOS-DA-PARTICIPACAO-
EM-BANCA-DE-MESTRADO/@HOME-PAGE"/>'
);
INSERT INTO DETALHAMENTO_BANCA (COD_PARTICIPACAO,
NOME_CANDIDATO, NOME_INST, NOME_CURSO)
VALUES (codparticipacao,
'<xsl:value-of select = "DETALHAMENTO-DA-PARTICIPACAO
-EM-BANCA-DE-MESTRADO/@NOME-DO-CANDIDATO"/>',
'<xsl:value-of select = "DETALHAMENTO-DA-PARTICIPACAO
-EM-BANCA-DE-MESTRADO/@NOME-INSTUICAO"/>',
'<xsl:value-of select = "DETALHAMENTO-DA-PARTICIPACAO
-EM-BANCA-DE-MESTRADO/@NOME-CURSO"/>'
);
<xsl:for-each select = "PARTICIPANTE-BANCA»

```

```

INSERT INTO PARTICIPANTE_BANCA (COD_PARTICIPACAO,
NOME_PARTICIPANTE, NOME_CITACAO, ORDEM_PARTICIPANTE)
VALUES (codparticipacao,
'<xsl:value-of select="@NOME-COMPLETO-DO-PARTICIPANTE
-DA-BANCA"/>',
'<xsl:value-of select="@NOME-PARA-CITACAO-DO-PARTICIPANTE
-DA-BANCA"/>',
'<xsl:value-of select="@ORDEM-PARTICIPANTE"/>'
);
</xsl:for-each>
INSERT INTO INF_ADICIONAIS_BANCA (COD_PARTICIPACAO,
INFORMACOES)
VALUES (codparticipacao,
'<xsl:value-of select="INFORMACOES-ADICIONAIS/@DESCRICAO
-INFORMACOES-ADICIONAIS"/>'
);
</xsl:template>
END;
</xsl:stylesheet>

```