OLINDA NOGUEIRA PAES CARDOSO

GESTÃO DO CONHECIMENTO USANDO DATA MINING: ESTUDO DE CASO NA UFLA

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-Graduação em Administração, área de concentração em Organizações, Estratégias e Gestão, para a obtenção do titulo de "Mestre".

Orientadora
Profa. Dra. Rosa Teresa Moreira Machado

LAVRAS
MINAS GERAIS – BRASIL
2005

Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da Biblioteca Central da UFLA

Cardoso, Olinda Nogueira Paes
Gestão do conhecimento usando Data Mining: estudo de caso na UFLA / Olinda Nogueira Paes Cardoso. -- Lavras : UFLA, 2005.
124 p. : il.

Orientador: Rosa Teresa Moreira Machado.
Dissertação (Mestrado) – UFLA.
Bibliografia.

1. Gestão do conhecimento. 2. DCBD. 3. Data Mining. 4. Plataforma Latttes. I. Universidade Federal de Lavras. II. Título.

CDD-658.4038

OLINDA NOGUEIRA PAES CARDOSO

GESTÃO DO CONHECIMENTO USANDO DATA MINING: ESTUDO DE CASO NA UFLA

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-Graduação em Administração, área de concentração em Organizações, Estratégias e Gestão, para a obtenção do titulo de "Mestre".

APROVADA em 27 de janeiro de 2005

Prof. Dr. André Luiz Zambalde — UFLA

Profa. Dra. Fernanda Cláudia Alves Campos — UFJF

Profa. Dra. Rosa Teresa Moreira Machado
UFLA
(Orientadora)

LAVRAS
MINAS GERAIS – BRASIL

DEDICATÓRIA

Dedico este trabalho a toda a minha família, pois, mesmo distante, sempre que precisei, ela deu-me o apoio necessário para que eu continuasse lutando e concluísse com sucesso mais uma etapa da minha vida. Em especial, dedico a meus pais, por terem garantido, durante toda a vida, a educação necessária para que eu chegasse até este momento.

AGRADECIMENTOS

Agradeço primeiramente a Deus, pela presença fundamental, pela força e, principalmente, por amparar-me nos momentos difíceis.

Expresso também meu maior agradecimento e o meu profundo respeito aos meus colegas de trabalho do Departamento de Ciência da Computação da UFLA, em especial ao professor André Luiz Zambalde, que muito colaborou para a elaboração deste trabalho; aos professores do Departamento de Administração e Economia da UFLA, que conheci há tão pouco tempo, mas dos quais adquiri muito conhecimento e uma grande admiração e aos meus queridos alunos do curso de graduação em Ciência da Computação da UFLA que, durante todo o tempo, me deram o apoio de que eu precisava para que eu pudesse ter concluído este trabalho.

Por fim, um agradecimento especial à professora Rosa Teresa, que mesmo sem me conhecer bem, sempre teve confiança em meu trabalho, sempre me orientou com sabedoria e tanto se esforçou para que este trabalho fosse concluído.

SUMÁRIO

LISTA DE FIGURAS

Figura **Página**

RESUMO

CARDOSO, Olinda Nogueira Paes. **Gestão do conhecimento usando *Data Mining***: estudo de caso na UFLA. 2005. 124p. Dissertação (Mestrado em Administração) – Universidade Federal de Lavras, Lavras, MG.[1]

O conhecimento é um dos mais importantes recursos de uma organização, possibilitando ações inteligentes. O processo de gestão do conhecimento abrange toda a forma de gerar, armazenar, distribuir e utilizar o conhecimento, tornando necessária a utilização de tecnologias de informação para facilitar esse processo, devido ao grande aumento no volume de dados. Processar e analisar as informações geradas pelas enormes bases de dados atuais de forma correta são requisitos essenciais para uma boa tomada de decisão. Uma metodologia emergente que tenta solucionar este problema da análise de grandes quantidades de dados é a Descoberta de Conhecimento em Banco de Dados e o *Data Mining*, uma técnica que faz parte desta metodologia. As Instituições de Ensino Superior (IES) são organizações voltadas para o conhecimento. Levando em consideração os problemas enfrentados pelas IES com o gerenciamento dos dados, o presente trabalho tem como objetivo desenvolver, aplicar e analisar uma ferramenta de *Data Mining*, para extrair conhecimento referente à produção científica das pessoas envolvidas com a pesquisa na Universidade Federal de Lavras (UFLA). Para isso, foi criado um banco de dados a partir de arquivos extraídos da Plataforma Lattes. O referencial teórico relaciona-se com gestão do conhecimento, descoberta de conhecimento em bancos de dados, *Data Mining* e gestão de universidades. São abordados alguns pontos críticos da política de gestão da pesquisa científica na UFLA e, por fim, uma descrição da Plataforma Lattes. A metodologia utilizada envolveu a pesquisa bibliográfica, a pesquisa documental e o método do estudo de caso, uma vez que apenas foram utilizados dados referentes à produção científica da UFLA. As limitações encontradas na análise dos resultados indicam que ainda é preciso padronizar o modo do preenchimento dos currículos Lattes para refinar as análises e, com isso, estabelecer indicadores. A contribuição foi gerar um banco de dados estruturado, que faz parte de um processo maior de desenvolvimento de indicadores de ciência e tecnologia, com o objetivo de auxiliar na elaboração de novas políticas de gestão científica e tecnológica e aperfeiçoamento do sistema de ensino superior do país.

---

[1] **Orientadora:** Profa. Dra. Rosa Teresa Moreira Machado - UFLA

ABSTRACT

CARDOSO, Olinda Nogueira Paes. **Knowledge Management using Data Mining**: a case study in UFLA. 2005. 124p. Dissertation (Master Program in Administration) – Universidade Federal de Lavras, Lavras, MG. [2]

Knowledge is one of the most important resources of an organization, making possible intelligent actions. The process of administration of knowledge embraces every form of generation, storage, distribution and use of the knowledge, making necessary the use of information technologies to facilitate that process, due to the great increase in the volume of data. To process and analyze the information generated by the enormous bases of current data in a correct way is one of the essential requirements for a good decision. An emergent methodology that tries to solve this problem of the analysis of great amounts of data is the Knowledge Discovery in Database (KDD) and Data Mining, a technique that is part of this methodology. The Universities are organizations turned to the knowledge. Considering the problems faced by Universities with the administration of the data, the present work aims to develop, apply and analyze a tool of Data Mining, to extract knowledge regarding the people's scientific production involved with the research at the Federal University of Lavras (UFLA). Thus, a database was created generated from extracted files of the Lattes Platform. The theoretical referential is linked to the management of the knowledge, knowledge discovery in databases, Data Mining and the administration of Universities. Some critical points of the administration politics of the scientific research at UFLA are approached, and finally, a description of the Lattes Platform. The methodology used involved the bibliographical research, the documental research, and the method of the case study, once it was just used referring data to the scientific production of UFLA. The limitations found in the analysis of the results indicate that is still necessary to standardize the way to fill out the Lattes curricula, to refine the analyses and, establish indicators. The contribution was to generate a structured database, which is part of a larger process of development of science and technology indicators, with the objective of aiding the elaboration of new politics of scientific and technological management and improvement of the superior education system of the country.

---

2 Guidance: Profa. Dra. Rosa Teresa Moreira Machado – UFLA

1 INTRODUÇÃO

Nos dias atuais, pode-se observar que as organizações vêm sofrendo uma grande mudança na forma de gerir recursos materiais, pessoais e, principalmente, as 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 organizações, mais flexíveis e atuando em rede e à nova ordem geopolítica mundial, aberta e volátil.

É neste contexto que surge a chamada era da sociedade da informação e do conhecimento, em que a informação constitui a principal matéria-prima, o conhecimento é utilizado na agregação de valor a produtos e serviços e a tecnologia constitui um elemento vital para as mudanças, em especial o emprego da tecnologia sobre acervos de informação.

O conhecimento tem sido reconhecido como um dos mais importantes recursos de uma organização, tornando possíveis ações inteligentes nos planos organizacional e individual, induzindo a inovações e capacidade de continuamente criar produtos e serviços excelentes em termos de complexidade, flexibilidade e criatividade.

O processo de gestão do conhecimento abrange toda a forma de gerar, armazenar, distribuir e utilizar o conhecimento, tornando necessária a utilização de tecnologias de informação para facilitar esse processo, devido ao grande aumento no volume de dados.

Ao longo do tempo, percebeu-se que a velocidade de coleta de informações era muito maior do que a velocidade de processamento ou análise das mesmas. Isto gera um problema e uma contradição, pois as organizações, por possuírem uma grande quantidade de dados, possuem uma falsa sensação de que estão bem informadas; porém, estas informações de nada servem se não forem analisadas de forma correta e em tempo hábil.

Em outras palavras, a coleta e o armazenamento de dados, por si só, não contribuem para melhorar a estratégia da organização. É necessário que se façam análises sobre essa grande quantidade de dados, estabelecendo-se indicadores para descobrir padrões de comportamento implícitos nos dados, assim como relações de causa e efeito. Processar e analisar as informações geradas pelas enormes bases de dados atuais de forma correta estão entre os requisitos essenciais para uma boa tomada de decisão.

Num ambiente extremamente mutável como o das organizações na atualidade, torna-se necessária a aplicação de técnicas e ferramentas automáticas que agilizem o processo de extração de informações relevantes de grandes volumes de dados. Uma metodologia emergente, que tenta solucionar este problema da análise de grandes quantidades de dados e ultrapassa a habilidade e a capacidade humanas, é a descoberta de conhecimento em banco de dados.

Data Mining, ou mineração de dados, é uma técnica que faz parte de uma das etapas da descoberta de conhecimento em banco de dados. Ela é capaz de revelar, automaticamente, o conhecimento que está implícito em grandes quantidades de informações armazenadas nos bancos de dados de uma organização. Esta técnica pode fazer, dentre outras, uma análise antecipada dos eventos, possibilitando prever tendências e comportamentos futuros, permitindo aos gestores tomar decisões baseadas em fatos e não em suposições.

É possível extrair, por exemplo, um grande número de informações úteis a partir da análise da produção científica, tecnológica e bibliográfica desenvolvida na Universidade Federal de Lavras (UFLA). Para isso, foi criado um banco de dados gerado a partir de arquivos extraídos da Plataforma Lattes[3] e, posteriormente, foi desenvolvida uma ferramenta de *Data Mining*, utilizando os

[3] Conjunto de Sistemas de Informações, bases de dados e portais para internet voltados para a gestão de Ciência e Tecnologia (Grupo Stella, 2002a).

recursos de um sistema gerenciador de banco de dados, para identificar padrões e tendências, gerando base para a gestão do conhecimento na instituição.

As Instituições de Ensino Superior (IES) são organizações voltadas para o conhecimento. Ao longo dos últimos anos, diversos autores vêm discutindo como avaliar a qualidade dos serviços prestados por estas instituições e nunca se questionou tanto a qualidade e os valores cobrados por esses serviços. Tem-se acentuado a necessidade de se refletir sobre a gestão das IES, preparando-as para as transformações que estão ocorrendo no ambiente em que operam. Cabe às próprias IES gerar soluções para gestão de políticas de ciência, tecnologia e inovação, que tenham um horizonte maior de planejamento a partir dessa enorme massa de dados ainda subutilizados. Este trabalho é uma contribuição neste sentido.

Levando em consideração os problemas enfrentados pelas universidades com o gerenciamento dos dados, além de diversas limitações encontradas na gestão dos sistemas de informação, o presente trabalho utilizou a técnica de *Data Mining*, extraindo conhecimento e contribuindo para a melhoria do preenchimento dos dados na Plataforma Lattes.

Este trabalho é uma etapa do processo de desenvolvimento do conhecimento, que pode servir de apoio à tomada de decisão, possibilitando, no futuro, a criação de indicadores para efeito comparativo entre instituições de ensino superior e de apoio à gestão da política científica e tecnológica e aperfeiçoamento do sistema de ensino superior do país.

1.1 Objetivos e escopo do trabalho

Como parte do processo de descoberta de conhecimento em banco de dados, este trabalho tem como objetivo geral desenvolver, aplicar e analisar uma ferramenta de *Data Mining*, para extrair conhecimento referente à produção científica dos professores da UFLA.

Como objetivos específicos, têm-se:

1. selecionar e tratar os dados disponíveis na Plataforma Lattes referentes à pesquisa científica na UFLA;
2. implementar um programa para transformar os dados selecionados num banco de dados;
3. desenvolver uma ferramenta automática de descoberta de conhecimento, utilizando a técnica de *Data Mining* e descrevê-la;
4. descrever as informações geradas e analisá-las.

Esta dissertação está organizada como se segue: no Capítulo 2 é apresentado o referencial teórico necessário para o entendimento e realização do projeto, tais como: gestão do conhecimento, com seus objetivos e alguns procedimentos adotados; descoberta de conhecimento em bancos de dados; *Data Mining*, sua definição, funcionamento básico e suas principais etapas; gestão de universidades, seus principais aspectos. São abordados alguns pontos críticos da política de gestão da pesquisa científica na Universidade Federal de Lavras e, por fim, uma descrição da Plataforma Lattes e seus principais componentes.

O Capítulo 3 trata da metodologia adotada para a realização deste trabalho, bem como as atividades realizadas e ferramentas utilizadas para a sua viabilização. No Capítulo 4 são apresentados os resultados e algumas discussões sobre os mesmos.

As considerações finais, que foram retiradas de todo o processo de desenvolvimento deste projeto, estão sintetizadas no Capítulo 5, além de sugestões de continuidade do mesmo.

2 REFERENCIAL TEÓRICO

2.1 Gestão do conhecimento

De acordo com Tarapanoff (2001), as mudanças que vêm ocorrendo nas organizações, atualmente, convergem para a quebra de um paradigma histórico e, por meio deste, entramos na era sociedade da informação e do conhecimento. A informação como principal matéria-prima das organizações é um insumo comparável à energia que alimenta um sistema; o conhecimento é utilizado na agregação de valor a produtos e serviços; a tecnologia constitui um elemento vital para as mudanças, em especial o emprego da tecnologia sobre acervos de informação. 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, utilizando, para isso, seu conhecimento, que é considerado um dos mais importantes recursos de uma organização. O conceito de conhecimento, com base em inúmeras definições, envolve estruturas cognitivas que representam determinada realidade. Segundo Krogh et al. (2001), citados por Alvarenga et al. (2002), conhecimento é caracterizado como sendo uma crença verdadeira e justificada, significando que as pessoas interpretam as informações conforme sua visão de mundo, ou pode ser entendido como a experiência, o entendimento e o *know-how* prático que o ser humano possui e que guiam suas decisões e ações.

Neste cenário, a gestão do conhecimento emerge como a área que estuda como as organizações podem entender o que elas conhecem, o que elas necessitam conhecer e como elas podem tirar o máximo proveito do conhecimento (Carvalho, 2000). Como o processo de gestão do conhecimento é abrangente e complexo, torna-se necessária a utilização de tecnologias da

informação, principalmente no que se refere à análise da grande quantidade de informação que é armazenada.

A velocidade na ocorrência das mudanças em nossa sociedade e o aumento da competição dos mercados globais têm contribuído para o questionamento sobre quais seriam os pilares fundamentais do sucesso das organizações. Além disso, o ciclo de desenvolvimento de produtos e serviços nas organizações tem sido drasticamente reduzido e estas buscam cada vez mais qualidade, inovação e velocidade para permanecerem no mercado. Para sobreviver, as organizações precisam aprender a diferenciar seus produtos e serviços por meio 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 organizações. As corporações estão se diferenciando umas das outras não tanto pelo que produzem, mas pelo que sabem.

Nesta mesma linha de raciocínio, Barroso & Gomes (1999), também citados por Pirolla (2002), afirmam 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/organizacionais. Neste contexto, o desafio de utilizar o conhecimento que existe nas organizações 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-chave. Produtos ou serviços únicos ou produzidos a um custo menor do que os concorrentes dependem de um conhecimento superior sobre o processo de produção que gera um projeto superior.

Saber como fazer coisas de forma eficaz e eficiente e de modo que as outras organizações não possam copiar é uma das principais fontes 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 organização, afirmam Laudon & Jane (1999), ao citar em Favela (1997).

Antes de chegar a uma definição do que seja gerenciar o conhecimento, é necessário conceituar "conhecimento". Diversos autores (Adriaans & Zantinge, 1996; Fayyad et al., 1996; Elmasri & Navathe, 2002; Navega, 2002; Amo, 2003; Moxton, 2004) fazem uma distinção ascendente entre os termos dado, informação e conhecimento, que poderiam ser sintetizados como se segue.

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.

2.1.1 Dos dados à sabedoria

Assim como um organismo vivo, as organizações recebem informação do meio ambiente e também atuam sobre ele. Segundo Navega (2002), durante essas atividades, é necessário distinguir vários níveis de informação. A Figura

2.1 apresenta um diagrama com a tradicional pirâmide da informação, na qual se pode notar o natural aumento de abstração conforme se sobe de nível.

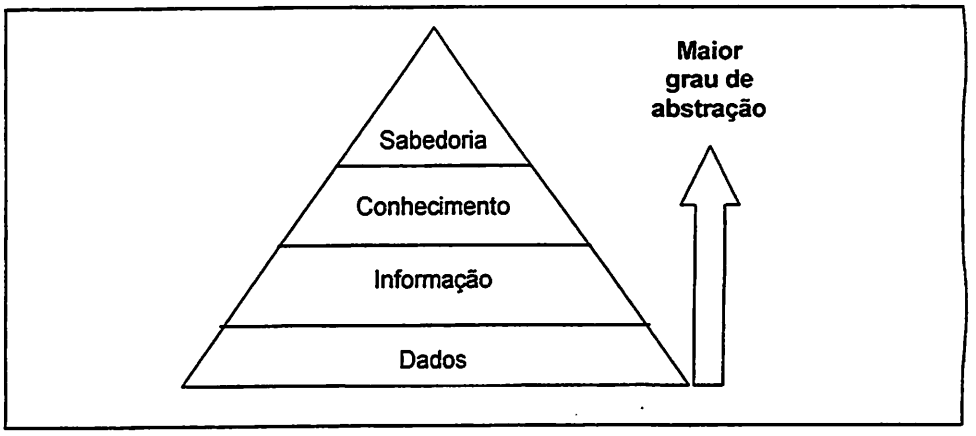


FIGURA 2.1 Pirâmide da informação
Fonte: Navega (2002, p. 3).

Traduzido para uma organização atual e considerando a adoção e uso de novas tecnologias de informação e comunicação, esse diagrama fica como apresentado na Figura 2.2.

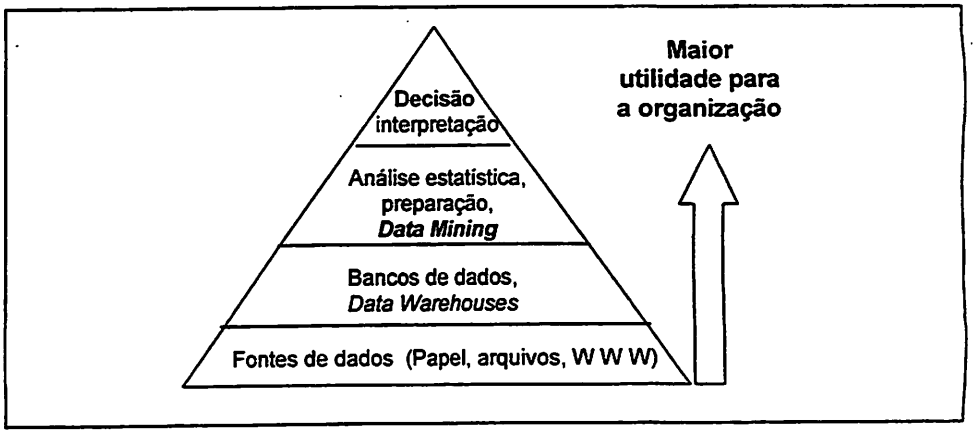


FIGURA 2.2 Pirâmide da evolução do valor estratégico da informação nas organizações
Fonte: Elaborado pela autora baseado em Navega (2002) e em Quoniam (2001).

O fundamental a se perceber neste diagrama é a sensível redução de volume de dados que ocorre cada vez que se sobe de nível. Essa redução de volume é uma natural conseqüência do processo de abstração. Abstrair, no sentido que aqui está sendo utilizado, é representar uma informação por meio de correspondentes simbólicos e genéricos. A importância disto é perceber que, para ser genérico, é necessário "perder" um pouco dos dados, para só conservar a "essência" da informação. O processo de *Data Mining* localiza padrões por meio da judiciosa aplicação de processos de generalização, algo que é conhecido como indução (Amo, 2004). Na Seção 2.3 este processo será mais detalhado.

2.1.2 Tipos de conhecimento

Segundo Tarapanoff (2001), o conhecimento organizacional pode ser classificado em dois tipos. O primeiro 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).

Considera-se 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 organização torna-se uma responsabilidade gerencial central. O processo de, ativa e sistematicamente, gerenciar e alavancar o armazenamento de conhecimento numa organização é chamado de gestão do conhecimento (Laudon & Jane, 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).

Malhotra (1998), citado por Parrini (2002), 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, ela abrange processos organizacionais que buscam a combinação sinérgica de dados e a 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 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), citado 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 ela 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, há idéias básicas que permeiam todos eles, permitindo a definição de um modelo genérico. Tarapanoff (2001) sugere sete processos na composição do modelo genérico de gestão do conhecimento:

- identificação - reconhecer que competências são críticas para o melhor desempenho da organização (competências essenciais);
- captura - adquirir conhecimentos, habilidades e experiências necessárias para criar e manter as competências essenciais e áreas de conhecimento selecionadas e mapeadas;
- seleção e validação - filtrar o conhecimento, avaliar sua qualidade e sintetizá-lo para fins de aplicação futura (fortemente associada ao processo de captura);
- organização e armazenagem - refletir sobre algumas questões consideradas básicas, tais como: que conhecimento a organização quer ou deve guardar; de que conhecimento a organização necessita; que conhecimento deve ser ignorado ou descartado, e qual a melhor forma de recuperar o conhecimento;

- 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;
- 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;
- criação do conhecimento - envolve as seguintes dimensões: aprendizagem, externalização do conhecimento, lições aprendidas, pensamento criativo, pesquisa, experimentação, descoberta e inovação.

Neste trabalho, devido à sua principal característica, será dada uma maior atenção ao último processo de criação do conhecimento.

2.1.3 Criando conhecimento

De acordo com Tarapanoff (2001), 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, acontecem por meio de 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 2.3 apresenta uma ilustração desses quatro processos:

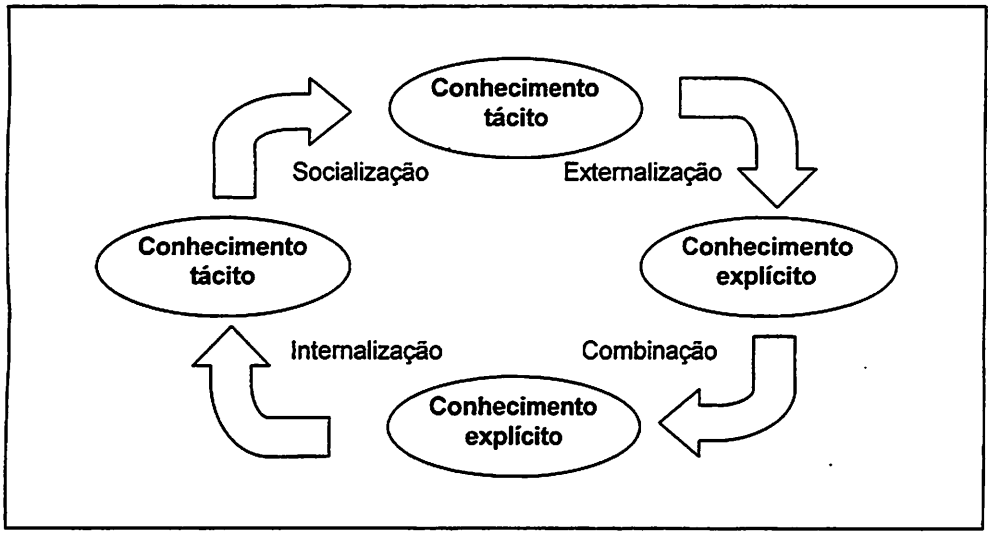


FIGURA 2.3 Os quatro processos de conversão do conhecimento
Fonte: Tarapanoff (2001, p. 136)

- 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.

Para a 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 que as organizações possuem. 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 necessidade de informações disponíveis vem crescendo assustadoramente nos últimos anos e vários fatores contribuíram para este incrível aumento. O baixo custo de armazenagem pode ser visto 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 conseqüência, 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 à tomada 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 gestores responsáveis pela tomada de decisão uma atenção redobrada aos ambientes interno e externo da organização. Muitas vezes, más decisões são definidas, não pela inexistência do conhecimento para se escolher

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 do inglês *Knowledge Discovery in Databases*), processo este que possui o *Data Mining* como principal etapa (Amo, 2003). Ou seja, para que o conhecimento seja descoberto, técnicas de *Data Mining* (mineração de dados) devem ser aplicadas. A Figura 2.4 representa um esquema simplificado deste processo.

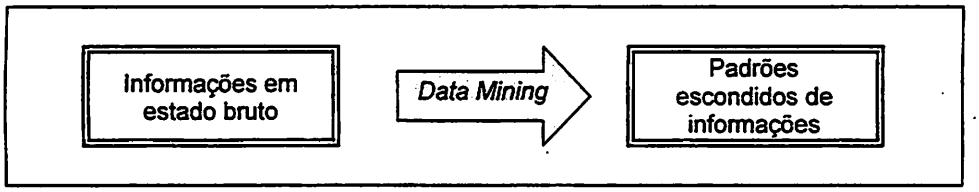


FIGURA 2.4 Esquema simplificado do processo de Data Mining
Fonte: Quoniam et al. (2001, p. 21)

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. Em outras palavras, a mineração de dados seria uma etapa do processo de DCBD.

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 o termo

Data Mining refere-se à aplicação de algoritmos[4] para a extração de padrões em um conjunto de dados, mas sem os passos adicionais em um processo de descobrimento do conhecimento.

Segundo Navega (2002), vale ressaltar que encontrar padrões requer que os dados brutos sejam sistematicamente simplificados, de forma a desconsiderar aquilo que é específico e privilegiar aquilo que é genérico. Faz-se isso porque não parece haver muito conhecimento a extrair de eventos isolados.

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.

Segundo Santos (2002), 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, por meio de uma consulta,

[4] Algoritmo é uma seqüência de passos definidos numa linguagem de programação, destinados a resolver um problema ou atingir algum objetivo específico.

um conjunto de dados para sua análise. Após a geração desse conjunto de dados, são utilizadas ferramentas de análises e visualização.

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 interpretação de resultados. A seguir, cada uma dessas etapas será observada com maior detalhe. A Figura 2.4 representa estas etapas do processo de DCBD.

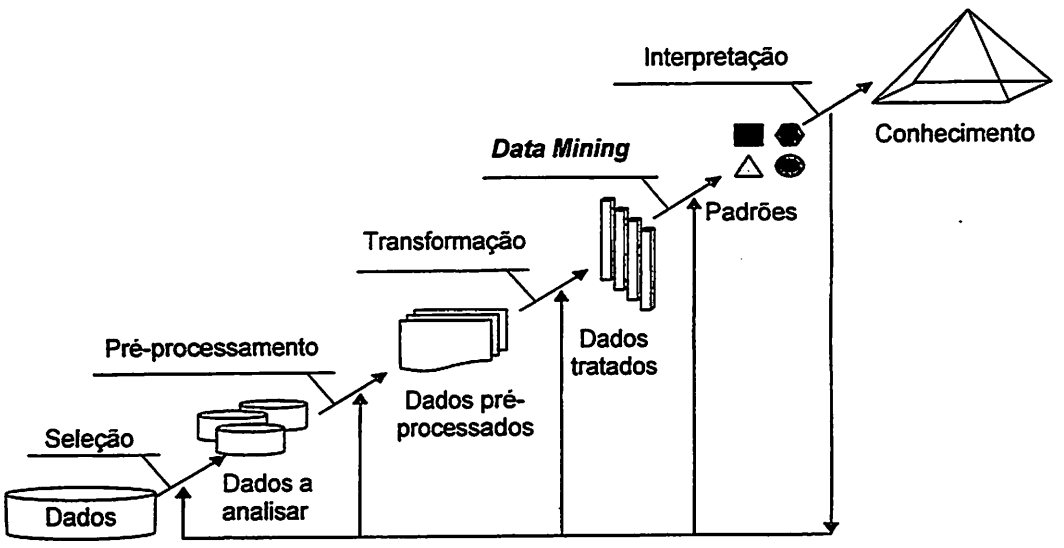


FIGURA 2.5 Processo de descoberta de conhecimento em banco de dados
Fonte: Santos (2002, p. 9).

2.2.1 Seleção ou filtragem dos dados

Segundo Navega (2002), as bases de dados são, na maioria das vezes, dinâmicas, incompletas, redundantes, ruidosas e esparsas, ou seja, contêm muitos erros, arquivos repetidos ou com dados nulos que devem ser excluídos.

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. É uma tarefa que exige um conhecimento tácito muito grande do analista, pois ele precisa discernir entre um dado realmente incorreto e uma exceção (*outlier*) à regularidade do conjunto de dados que deve ser mantida no banco de dados.

Segundo Fayyad et al. (1996), 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 mais de uma vez e inconsistência no domínio é um erro que pode ser gerado na entrada de dados no arquivo. Além disso, o problema de filtragem dos dados pode ser dinâmico, ou seja, se o processo de DCBD for utilizar dados coletados continuamente e houver 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 (Morzy et al., 2000).

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, por meio de um conjunto de testes, em que 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.

Ainda segundo Morzy et al. (2000), as ferramentas de análise podem ser baseadas em algoritmos ou em visualização. No primeiro caso, um modelo é especificado por meio da associação de variáveis de entrada (independente) e variáveis de saída (dependente). No segundo caso, a hipótese é especificada pela visualização do conjunto de dados esperados como resultado e da seleção de elementos nos dados. A própria visualização produzida é o modelo e seu poder explicativo pode ser observado por meio da visualização.

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 à formulação das hipóteses. Nessa fase, os três principais passos do processo 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 agrupamento (*clustering*). Para a seleção do modelo de *Data Mining*, uma grande variedade de

modelos de análises podem ser utilizada, tais 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 (Santos, 2002).

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 objeto de estudo.

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, seguem as quatro etapas descritas anteriormente, com a utilização de diversas ferramentas auxiliares. Por fim, após a geração dos resultados, tem-se como produto 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.

Por meio dos pontos observados anteriormente, pode-se 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

Talvez a definição mais importante de *Data Mining* tenha sido elaborada por Fayyad et al. (1996, p. 4): "...o processo não-trivial de identificar, em dados, padrões válidos, novos, potencialmente úteis e ultimamente compreensíveis".

Data Mining, ou Mineração de Dados, é uma área de pesquisa multidisciplinar, incluindo principalmente as tecnologias de bancos de dados, inteligência artificial, estatística, reconhecimento de padrões, sistemas baseados em conhecimento, recuperação da informação, computação de alto desempenho e visualização de dados. 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 de diversas áreas e cada área possui abordagens específicas, adequadas para as suas necessidades.

A Mineração de Dados na área da computação teve início nos anos 1980, quando os profissionais das organizações começaram a se preocupar com os grandes volumes de dados informáticos estocados e inutilizados. Naquela época, segundo Amo (2004), *Data Mining* consistia essencialmente em extrair informação de gigantescas bases de dados da maneira mais automatizada possível. Atualmente, *Data Mining* consiste, sobretudo, na análise dos dados após a extração.

Os seguintes pontos são algumas das razões pelas quais o *Data Mining* vem se tornando necessário para uma boa gestão organizacional: (a) os volumes de dados são muito importantes para um tratamento utilizando somente técnicas clássicas de análise, (b) o usuário final não é necessariamente um estatístico e (c) a intensificação do tráfego de dados (navegação na internet, catálogos *on-line* etc.) aumenta a possibilidade de acesso aos dados.

Segundo Oracle (2004), no grande mercado competitivo, um dos fatores críticos para as organizações é o gerenciamento dos seus bens mais valiosos – seus clientes e as informações que elas têm sobre eles 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 organização, como, por exemplo, entender o comportamento dos seus funcionários e atuar de acordo com as peculiaridades de cada um.

Segundo Amo (2003), vale ressaltar que é importante distinguir o que é uma tarefa e o que é uma técnica de mineração de dados. A tarefa consiste na especificação do "que" se deseja buscar nos dados, que tipo de regularidades ou categorias de padrões tem-se interesse em encontrar, ou que tipo de padrões poderiam surpreender. Já a técnica de mineração consiste na especificação de métodos que garantam "como" descobrir os padrões que interessam. Dentre as principais técnicas utilizadas em mineração de dados, estão técnicas estatísticas e técnicas de inteligência artificial.

Segundo King (2003), *Data Mining* é um modo de procurar relações interessantes escondidas em um grande conjunto de dados, tais como padrões de *clustering* (agrupamentos) e aproximações de funções. 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 contêm erros. Ao examinar os dados, um algoritmo pode explorar milhares de prováveis regras, utilizando diversas técnicas para escolher entre elas.

Decker & Focardi (1995) 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.

Grossman et al. (2002) definem *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 por meio dos dados. Por outro lado, as técnicas de *Data Mining* são baseadas na descoberta, na medida em que os padrões são automaticamente extraídos do conjunto de dados.

De acordo com Moxton (2004), *Data Mining* é um conjunto de técnicas utilizadas para explorar exaustivamente e trazer à superfície relações complexas em um conjunto grande de dados. Uma diferença significante entre as técnicas de *Data Mining* e outras ferramentas analíticas é a abordagem utilizada para explorar as inter-relações entre os dados, semelhante à abordagem dada por Grossman et al. (2002), que também diferenciam as técnicas de *Data Mining* com relação às técnicas analíticas entre a abordagem de suposição e a abordagem de descoberta. Segundo esses autores, discordando de outros pesquisadores, as técnicas de *Data Mining* não pressupõem que as relações entre os dados devam ser conhecidas *a priori*. Isto significa que, ao ser aplicada a técnica de *Data Mining*, novas relações entre os dados irão surgir.

Em outras palavras, o *Data Mining* nada mais é do que um conjunto de algoritmos matemáticos utilizado para produzir conhecimento analisando dados,

descobrindo tendências 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*[5] , por exemplo).

Algumas das grandes perguntas que os gestores gostariam de 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 organização? 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 a essas perguntas que as organizações estão utilizando o *Data Mining*, pois 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 *insigths* e elementos para conhecimento tácito.

Segundo Amo (2004), existem diversas medidas objetivas para avaliar o grau de interesse que um padrão pode apresentar ao usuário. Tais medidas são baseadas na estrutura do padrão descoberto e em estatísticas apropriadas. Por

---

[5] *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 à gestão (Elmasri e Navathe, 2002).

exemplo, uma medida objetiva para avaliar o interesse de uma regra de associação é o suporte, representando a porcentagem de transações de um banco de dados de transações onde a regra se verifica. Em geral, cada medida objetiva está associada a um limite mínimo de aceitação, que pode ser controlado pelo usuário.

Além das medidas objetivas, o usuário pode especificar medidas subjetivas para guiar o processo de descoberta, refletindo suas necessidades particulares. Afinal, padrões que são interessantes, segundo medidas objetivas, podem representar conhecimento óbvio e, portanto, sem interesse. Pode-se, por exemplo, medir o grau de interesse de um padrão pelo fato de ele ser inesperado pelo usuário.

Medidas (objetivas ou subjetivas) de avaliação do grau de interesse por padrões são essenciais para a eficiência do processo de descoberta de padrões. Tais medidas podem ser usadas durante o processo de mineração ou após o processo, a fim de classificar os padrões encontrados de acordo com o interesse de um dado usuário, filtrando e eliminando os não interessantes. Em termos de eficiência, é importante incorporar medidas de interesse que restrinjam o espaço de busca dos padrões durante o processo de descoberta, e não após o processo ter terminado.

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 organização que emprega a técnica de *Data Mining* é capaz de: criar parâmetros para entender o comportamento dos dados, que podem ser referentes a pessoas envolvidas com a organização; identificar afinidades entre dados que podem ser, por exemplo, entre pessoas e

produtos e ou serviços; prever hábitos ou comportamentos das pessoas e analisar hábitos para se detectar comportamentos fora do padrão; dentre outros.

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

- previsão - 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;
- classificação - pode repartir os dados de modo que diferentes classes ou categorias possam ser identificadas com base em combinações de parâmetros;
- 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), Elmasri & Navathe (2002) e Amo (2003), o conhecimento descoberto durante a fase de *Data Mining* pode ser descrito de acordo com cinco tarefas:

1. **Análise de regras de associação** – uma regra de associação é um padrão da forma $X \rightarrow Y$, em que X e Y são conjuntos de valores, ou seja, encontrar itens que determinem a presença de outros em uma mesma transação e estabelecer regras que correlacionam a presença de um conjunto de itens com um outro intervalo de valores para um outro conjunto de variáveis. Exemplo: sempre que se orienta um aluno de doutorado, é publicado algum documento; descobrir regras de associação entre alunos de doutorado e número de publicações pode ser útil para melhorar a distribuição de orientados por professor.

2. **Classificação e predição** – classificação é o processo de criar modelos (funções) que descrevem e distinguem classes ou conceitos, baseado em dados conhecidos, com o propósito de utilizar este modelo para predizer a classe de objetos que ainda não foram classificados. O modelo construído baseia-se na análise prévia de um conjunto de dados de amostragem ou dados de treinamento, contendo objetos corretamente classificados. Exemplo: grupos de pesquisas já definidos contendo alguns professores e, a partir da análise de dados das pesquisas de outros professores que não pertencem a estes grupos, sugerir a sua entrada.
3. **Análise de padrões seqüenciais** – um padrão seqüencial é uma expressão da forma $<I_1, ...I_n>$, em que cada I_i é um conjunto de itens. A ordem em que estão alinhados estes conjuntos reflete a ordem cronológica em que aconteceram os fatos representados por estes conjuntos. Encontrar padrões ou comportamento previsível em um período de tempo significa que um comportamento particular em um dado momento pode ter como conseqüência outro comportamento ou seqüência de comportamentos dentro de um mesmo período de tempo. Exemplo: uma pessoa que cursou mestrado provavelmente cursará doutorado em um certo período de tempo.
4. **Análise de *clusters* (agrupamentos)** – diferentemente da classificação e predição, em que os dados estão previamente classificados, a análise de *clusters* trabalha sobre dados nos quais as classes não estão definidas. A tarefa consiste em identificar novos agrupamentos, que contenham características similares e agrupar os registros, ou seja, particionar (segmentar) uma dada população de eventos ou itens em conjuntos. Exemplo: professores de departamentos diferentes, que trabalham em grupos de pesquisas diferentes, poderiam estar trabalhando com o mesmo objeto e, dessa forma, seria sugerida a formação de um novo

agrupamento destas pessoas, podendo surgir assim um novo grupo de pesquisa ou reclassificá-lo.

5. **Análise de *outliers*** – um banco de dados pode conter dados que não apresentam o comportamento geral da maioria. Estes dados são denominados *outliers* (exceções). Muitos métodos de mineração descartam estes *outliers* como sendo ruído indesejado. Entretanto, em algumas aplicações, estes eventos raros podem ser mais interessantes do que eventos que ocorrem regularmente. Exemplo: descobrir padrões de comportamento de professores que publicam um número muito grande de artigos e que fogem ao padrão dos demais professores.

O *Data Mining* usa ferramentas de análise estatística, assim como técnicas da área de inteligência artificial, 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[6].

2.4 Informações sobre gestão de ciência, tecnologia e inovação e sua importância

A gestão de Ciência, Tecnologia e Inovação (CT&I) diz respeito à administração e desenvolvimento de estratégias e instrumentos organizacionais, envolvendo aspectos estruturais, culturais, políticos, tecnológicos, gerenciais e de serviços, de forma a promover a pesquisa viável e relevante (Hayashi et al., 2004).

A tomada de decisões no campo da CT&I é uma tarefa complexa, que tem sido simplificada a partir do desenvolvimento de indicadores de Ciência e

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

Tecnologia (C&T), propostos como ferramentas para auxiliar no planejamento, monitoramento e avaliação de resultados científicos das nações.

Hayashi (2002) afirma que analisar atividades de CT&I é um desafio para a definição de políticas públicas. O avanço do conhecimento produzido por pesquisadores tem de ser transformado em informação acessível para a sociedade, o que coloca os indicadores das atividades de CT&I no centro dos debates.

Na gestão de C&T devem ser consideradas a escolha de linhas de pesquisa prioritárias quanto à relevância para o desenvolvimento sócio-econômico e cultural, a execução mais eficiente das pesquisas e a conversão mais rápida de resultados obtidos em contribuições para a comunidade. Tais aspectos devem ser considerados em três níveis de gestão: o das políticas públicas, o institucional (universidades, institutos de pesquisa, empresas etc.) e o de programas e projetos específicos de pesquisa (Coelho, 2002).

No entanto, Hayashi (2002) afirma que as principais questões envolvidas neste âmbito dizem respeito à caracterização e construção de indicadores que devem ser discutidos e analisados a partir do contexto de produção das atividades científicas, sem deixar de considerar as limitações e dificuldades para o seu desenvolvimento. O objetivo do trabalho desta autora foi desenvolver uma metodologia de produção de indicadores para a análise de atividades de CT&I na Universidade Federal de São Carlos (UFSCar). Dessa forma, tais indicadores podem constituir instrumentos para a definição de políticas de C&T nas instituições federais de ensino superior, uma vez que retratam a estrutura, a situação e a performance das atividades de pesquisa científica e tecnológica, tanto para reprodução e geração de conhecimentos, como para criação de novos produtos e processos.

A sua metodologia inclui: a) revisão de literatura em CT&I e sociedade da informação, b) caracterização do local, c) coleta de dados na Plataforma

Lattes e d) produção de indicadores de CT&I do local, com o auxílio de ferramentas estatísticas automatizadas (Hayashi, 2002).

A pesquisa desta autora indica que, se acompanhados ao longo dos anos, os indicadores de C&T permitirão às instituições: desenvolver mecanismos para planejar monitorar e avaliar as atividades de pesquisa institucional; estabelecer diretrizes para o desenvolvimento de uma política de C&T sintonizada com os avanços do conhecimento na sociedade da informação; servir de instrumento para conhecimento do perfil do pesquisador, dos programas de pós-graduação e dos grupos de pesquisas institucionais; estabelecer critérios sobre a alocação de recursos humanos, físicos, de equipamentos e material, financeiros e orçamentários, disponíveis e ou mobilizados pela instituição; preservar a memória da atividade científica e tecnológica desenvolvida na instituição; analisar os padrões de publicação científica e tecnológica da instituição; fortalecer e direcionar as ações de organismos de fomento à pós-graduação e pesquisa, dentre outros (Hayashi, 2002)

2.4.1 Indicadores de ciência, tecnologia e inovação

A elaboração de indicadores de ciência e tecnologia tem sido sustentada por marcos teóricos provenientes de várias disciplinas. Polcuch (1999), citado em Hayashi et al. (2004), apresenta o modelo linear de inovação. Em sua argumentação, este autor assinala que, habitualmente, os trabalhos acerca de indicadores de C&T os consideram como sistemas, que se nutrem de insumos (*inputs*) e produz em produtos (*outputs*).

Este modelo linear tem sido utilizado para explicar o vínculo entre conhecimento e desempenho econômico e, a partir dele, os governos começaram a articular políticas públicas em relação à ciência. Esta visão deu origem ao "Modelo Linear de C&T" ou "Modelo Linear de Inovação", desenhado a partir de dois aforismos: a) a pesquisa básica (o conhecimento geral e um entendimento da natureza e de suas leis) deve ser conduzida sem a preocupação

com fins práticos; b) a pesquisa aplicada – converte as descobertas da pesquisa básica em inovações tecnológicas que vão ao encontro das necessidades da sociedade. Ao longo de vários anos, este modelo influenciou largamente universidades, porém, atualmente, vem sendo questionado.

Um novo modelo atribui às pesquisas duas coordenadas: uma, que dimensiona o avanço do conhecimento que a pesquisa propicia e outra que dimensiona a aplicação que dela decorre. Dessa maneira, uma pesquisa pode, ao mesmo tempo, contribuir significativamente para o avanço do conhecimento e ter grandes perspectivas de aplicações práticas.

Segundo Hayashi et al. (2004), existe uma relação entre a capacidade de produzir indicadores de C&T e a capacidade, por parte de governos e instituições do setor público e privado, de realizar investimentos em C&T. Nos últimos anos, o desenvolvimento de políticas e estratégias para execução de metas institucionais conduziu os organismos de ciência e tecnologia e setores públicos a elaborar instrumentos de medição que possibilitem uma gestão otimizada e racional de seus recursos.

A temática e a produção de indicadores de CT&I fazem parte da agenda científica de organismos e instituições, demonstrando a importância desse tema. O uso desses indicadores como subsídio para a construção de políticas em C&T, com foco na informação, é um dos exemplos da importância de trabalhos nessa área (Ferraz & Basso, 2003; Brisolla, 1998, citados em Hayashi *et al* (2004).

No contexto nacional, o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), criado em 1951, foi a primeira instituição que realizou esforços para gerar indicadores de C&T. Outras iniciativas de construção de indicadores provêm do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) do Ministério da Ciência e Tecnologia (MCT) e, no campo do ensino superior, da CAPES[7]. Segundo informações divulgadas pelo

[7] Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (www.capes.gov.br)

MCT em seu *site*, este Ministério passou a assumir, de forma centralizada, a responsabilidade pela organização e divulgação das informações de C&T do país.

Os indicadores construídos pelo MCT passaram por duas fases: no início, concentravam-se no que passou a se denominar indicadores de insumo, isto é, no dimensionamento dos recursos financeiros e humanos investidos em C&T. A mensuração se limitava à identificação dos recursos aplicados à pesquisa, permitindo a construção do que se chamou "Dispêndio Interno em P&D", e aos recursos humanos – e sua capacitação – dedicados a tais atividades. Estes indicadores de insumo, seguindo a tendência daqueles dos demais países, possuem as séries mais longas e detalhadas (Brasil, 2001).

Como menciona o MCT na apresentação dos Indicadores de C&T, tradicionalmente estes indicadores de insumo são desagregados segundo três dimensões:

1. a natureza da pesquisa: básica, aplicada e atividades científicas e técnicas correlatas;
2. os setores que executam ou financiam estas atividades: governo, instituições de ensino superior e empresas;
3. a classificação dos recursos de cada um destes setores, obedecendo a critérios específicos para o governo (segundo objetivos sócio-econômicos), as instituições de ensino superior (segundo áreas de conhecimento) e as empresas (segundo setores de atividade econômica).

Mais recentemente, foram desenvolvidos os chamados indicadores de resultados, de início, limitados à produção científica e, posteriormente, incorporados à produção de patentes e a transferência de tecnologia entre países.

A constituição e a implantação da Plataforma Lattes foram iniciativas conjuntas do MCT, CNPq, CAPES e FINEP[8]. A plataforma é integrada pelos sistemas Currículo Lattes e Diretório de Grupos de Pesquisa no Brasil, os quais apresentam a opção "Indicadores de Produção de C&T" e fornecem uma visão quantitativa dos itens de produção científica e tecnológica cadastrados no Currículo e Diretório, permitindo consultar as distribuições das diferentes variáveis cadastradas. A Plataforma Lattes, principal objeto de estudo deste trabalho, será detalhada na Seção 2.6.

Com relação aos dados do Diretório dos Grupos de Pesquisa no Brasil, foi organizada, pelo CNPq, uma hierarquização dos grupos de pesquisa vinculados às instituições de ensino superior, classificando-os em cinco estratos. Esta classificação tomou como indicador a densidade de pesquisadores qualificados por dois sistemas de avaliação já consagrados no Brasil, baseados em julgamentos desenvolvidos pelos próprios pares, de forma relativamente independente em relação ao Diretório entre si: o sistema de avaliação *ex-ante* dos projetos de pesquisa e dos currículos dos pesquisadores candidatos às bolsas de pesquisa concedidas pelo CNPq e o sistema de avaliação dos programas de pós-graduação empreendido pela CAPES (CNPq, 2004).

A hierarquização dos grupos de pesquisa realizada pelo CNPq coloca em evidência as concentrações geográfica e institucional da pesquisa desenvolvida no âmbito das IES; ordena as instituições sob a ótica da pesquisa científica por Grande Área de Conhecimento, tendo em conta os quantitativos de grupos de pesquisa classificados nos diferentes estratos, em termos absolutos e relativos e,

[8] A Financiadora de Estudos e Projetos é uma empresa pública vinculada ao Ministério da Ciência e Tecnologia – MCT, criada em 1967 com a finalidade de financiar a implantação de programas de pós-graduação nas universidades brasileiras. Sua missão é promover e financiar a inovação e a pesquisa científica e tecnológica em empresas, universidades, institutos tecnológicos, centros de pesquisa e outras instituições públicas ou privadas, mobilizando recursos financeiros e integrando instrumentos para o desenvolvimento econômico e social do país. (www.finep.gov.br)

ao final, averigúa a existência de correlação entre o grau de qualificação e a produtividade técnico-científica de tais grupos. O indicador de produtividade considera a produção de C&T (artigos, livros e capítulos de livros publicados, produção tecnológica desenvolvida, teses e dissertações defendidas sob orientação de pesquisadores pertencentes ao grupo) dos pesquisadores doutores, cadastrada com o auxílio do Sistema de Currículo Lattes.

Segundo Macias-Chapula (1998), o foco da produção de indicadores de CT&I esteve, por muitos anos, voltado para a medição dos insumos e, apenas mais recentemente, aumentou o interesse em medir os resultados das atividades científicas e tecnológicas. A produção de indicadores também tem se concentrado em âmbito nacional, institucional ou com enfoque em áreas do conhecimento específicas e ainda são escassos os indicadores das atividades de CT&I em níveis regionais e locais. Ainda segundo estes autores, esta é uma lacuna que precisa ser preenchida.

A partir desta realidade, Hayashi (2002) optou por construir os indicadores de produção científica institucionais, divididos basicamente em dois grupos: os indicadores de produção científica e tecnológica associada à pós-graduação (que envolve as produções caracterizadas como bibliográficas, as formas de divulgação restrita da produção científica e trabalhos publicados em eventos científicos, dentre outros) e os indicadores de produção científica e tecnológica associada aos grupos de pesquisa (além da produção bibliográfica, inclui a produção técnica e as orientações concluídas). Os indicadores de C&T produzidos para a UFSCar estão consolidados em um conjunto de 61 tabelas e 13 figuras e estão disponíveis para consulta *on-line* no endereço http://www.propg.ufscar.br/publica/indicador.pdf. Por se tratar de um estudo de caso realizado na UFSCar, sua replicação em outras instituições é limitada, mas pode servir como base para outros trabalhos.

2.4.2 Gestão de universidades

A gestão de uma instituição de ensino típica é formada por um conjunto de decisões assumidas a fim de obter um equilíbrio dinâmico entre missão, objetivos, meios e atividades acadêmicas e administrativas (Tachizawa & Andrade, 2002). O trabalho destes autores visa estabelecer um modelo de gestão aplicável às instituições de ensino superior (IES).

Segundo Alvarenga et al. (2002), o foco da gestão estratégica do conhecimento em IES está pautado em: 1) diferenciação: que busca diferenciar os produtos produzidos e os serviços ofertados pela organização, visando criar algo que seja considerado único no setor de atuação; 2) concentração: significa a capacidade de satisfazer o público-alvo, mas exige o estabelecimento de uma política funcional voltada para este segmento.

O trabalho desenvolvido por Alvarenga et al. (2002) apresenta uma visão do modelo de gestão do conhecimento proposto para ensino e pesquisa na Universidade Católica de Brasília (UCB), elaborado para suportar necessidades de informação e orientar o processo de gestão das atividades da Universidade, por meio da administração do conhecimento gerado internamente. Tem por propósito apropriar conhecimento, disseminá-lo e garantir sua incorporação aos serviços e processos de decisão, com foco no desenvolvimento humano.

A compreensão da instituição de ensino e da sua inter-relação com os demais agentes do ramo de atividades, o setor educacional ao qual pertence, é essencial para se desenvolver uma proposta de ferramenta de auxílio à gestão do conhecimento, objetivo geral deste trabalho. Faz-se necessário analisar finalidades 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), consideram a universidade uma organização prestadora de serviço que oferece produtos, que são os profissionais formados, capazes de se inserir no âmbito de trabalho e na sociedade em geral.

Vale ressaltar que cada instituição do sistema deve acoplar-se ao nível de gestão das políticas públicas e, para que isso ocorra, é necessário que cada instituição defina uma política própria e clara quanto a projetos científica e tecnologicamente viáveis e relevantes (Hayashi et al., 2004). Para isso, é necessário identificar suas capacidades específicas e combinações de recursos e competências, aproveitando bem suas características próprias, além de contextuais e estabelecer formas de parcerias com outras instituições do sistema de C&T.

Por parceiros, entendem-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 parceiro (colaborador ou fornecedor) 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 é feita diretamente para os clientes ou por meio de intermediários;
- **fornecedor** – faz-se o mapeamento dos professores existentes no mercado que, potencialmente, sejam úteis à instituição de ensino;
- **concorrente** – procura-se identificar sua origem, características, pontos fortes e pontos fracos;
- **ramo de atividades** – identifica-se a qual tipo de setor econômico 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.

2.4.3 Gestão do conhecimento nas relações universidade *x* empresa: prioridades distintas

Apesar da existência de uma analogia entre universidades e organizações mercadológicas (empresas), 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 em 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 com pesquisa e desenvolvimento, 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 (Alvarenga et al., 2002).

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, quando estabelecem parcerias ou convênios com universidades, 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.

Segundo Alvarenga (2002), a pesquisa é reconhecidamente o componente mais importante para consolidar um sistema de pós-graduação que, por sua vez, é o principal elemento levado em conta nas avaliações pelas quais passam as IES. Por sua natureza extremamente dinâmica, a pós-graduação exige grandes esforços de planejamento, gestão e articulação. Para tanto, segundo

Nonaka (1995), citado por Alvarenga et al. (2002), é necessário criar um ambiente que envolve cinco processos:

> *a) Compartilhamento do conhecimento tácito: ambiente onde o conhecimento pode ser explicitado e compartilhado entre os interessados; b) Criação dos conceitos: requer entendimento do grupo acerca do conhecimento, com base no foco e interesse da instituição ou de um grupo; c) Justificação dos conceitos: novos conceitos assumidos e entendidos como capazes de proporcionar ganhos pela agregação de valor aos produtos em decorrência da aquisição de conhecimento; d) Construção dos protótipos: novos processos gerados a partir do conhecimento incorporado, considerando seu real benefício; e e) Nivelamento do conhecimento: processo de difusão do conhecimento adquirido através da informação de maneira que possa ser transformado novamente em conhecimento* (Alvarenga et al. (2002), p. 3).

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. Além disso, foi aplicado todo o processo de descoberta de conhecimento em bancos de dados.

> *A pesquisa Bibliográfica é a que se desenvolve tentando explicar um problema, utilizando o conhecimento disponível a partir das teorias publicadas em livros ou obras congêneres. Na pesquisa bibliográfica o investigador levanta o conhecimento disponível na área, identificando as teorias produzidas, analisando-as e avaliando sua contribuição para descrever, compreender ou explicar o problema objeto de investigação. O objetivo da pesquisa bibliográfica, portanto, é o de conhecer e analisar as principais contribuições teóricas existentes sobre um determinado tema ou problema, tornando-se um instrumento indispensável para qualquer tipo de pesquisa* (Koche, 1997, p.122).

A pesquisa bibliográfica deu base a 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, teses e artigos científicos.

A pesquisa documental foi realizada em documentos referentes à pesquisa científica na UFLA, obtidos a partir do Lattes Extrator os quais proporcionaram informações úteis para as análises, as comparações e para o desenvolvimento da ferramenta de *Data Mining*. Também foram pesquisados

documentos da UFLA referentes às políticas de incentivo ao desenvolvimento de CT&I.

O método do estudo de caso é considerado um tipo de análise qualitativa. Não é uma técnica especifica; é um meio de organizar dados sociais preservando o caráter unitário do objeto social estudado (Goode & Hatt, 1969).

Bonoma (1985) coloca que o estudo de caso é uma descrição de uma situação gerencial. Este método, assim como os métodos qualitativos, são úteis quando o fenômeno a ser estudado é amplo e complexo, quando o corpo de conhecimentos existente é insuficiente para suportar a proposição de questões causais e nos casos em que o fenômeno não pode ser estudado fora do contexto onde naturalmente ocorre.

Yin (1989) afirma que o estudo de caso é uma inquirição empírica que investiga um fenômeno contemporâneo dentro de um contexto da vida real. De acordo com Yin (1989), a preferência pelo uso do estudo de caso deve ser dada quando do estudo de eventos contemporâneos, em situações nas quais os comportamentos relevantes não podem ser manipulados, mas é possível se fazer observações diretas e sistemáticas.

Os objetivos do método de estudo de caso, segundo Bressan (2000), são capturar o esquema de referência e a definição da situação de um dado participante, permitir um exame detalhado do processo organizacional e esclarecer aqueles fatores particulares ao caso que podem levar a um maior entendimento da causalidade.

Bonoma (1985) ao tratar dos objetivos da coleta de dados, coloca como objetivos do método do estudo de caso não a quantificação ou a enumeração, mas, em vez disso, a descrição, a classificação, o desenvolvimento teórico e o teste limitado da teoria. Em uma palavra, o objetivo é a compreensão.

De forma sintética, Yin (1989) apresenta quatro aplicações para o método do estudo de caso: (1) para explicar ligações causais nas intervenções na

vida real, (2) para descrever o contexto da vida real no qual a intervenção ocorreu, (3) para fazer uma avaliação, ainda que de forma descritiva, da intervenção realizada e (4) para explorar aquelas situações em que as intervenções avaliadas não possuam resultados claros e específicos.

Yin (1989) coloca que uma preocupação em relação a este método é o fato de ele fornecer pequena base para generalizações científicas, uma vez que, por estudar um ou alguns casos, não se constitui em amostra da população e, por isto, torna-se sem significado qualquer tentativa de generalização para populações.

Segundo Bressan (2000), o método do estudo de caso oferece significativas oportunidades para a gestão, pois pode possibilitar o estudo de inúmeros problemas de administração de difícil abordagem por outros métodos e pela dificuldade de se isolá-los de seu contexto na vida real.

O estudo de caso de que trata o presente trabalho foi realizado na Universidade Federal de Lavras (UFLA), mais especificamente nos setores envolvidos com o desenvolvimento de pesquisa científica. O estudo utilizou dados de fontes secundárias como base para as análises, extraídos dos currículos de pessoas ligadas, de forma direta e indireta, à pesquisa científica da UFLA. Estes dados foram disponibilizados pelo uso da ferramenta Lattes Extrator, que faz parte da Plataforma Lattes.

3.2 Procedimento metodológico

Foi realizada, inicialmente, uma pesquisa bibliográfica dos assuntos discutidos nas seções anteriores, com o objetivo de adquirir um embasamento teórico. Foi necessário conhecer o campo da descoberta de conhecimento em bancos de dados e *Data Mining*: processos, modelos, técnicas utilizadas, entre outros. Além disso, buscou-se conhecer os problemas enfrentados pelos gestores da Universidade Federal de Lavras, no que diz respeito às pesquisas científicas,

suas prioridades, metas, mecanismos e cultura, de modo geral. A pesquisa documental realizada na Pró-Reitoria de Pesquisa foi essencial para a realização deste trabalho.

Após a coleta de informações necessárias para o embasamento teórico, foi realizado um estudo documental nos arquivos extraídos da Plataforma Lattes. Por meio da interface *on-line* do Lattes Extrator (Figura 3.1), foram extraídos mais de mil currículos de professores, alunos, ex-alunos, mestrandos e doutorandos.

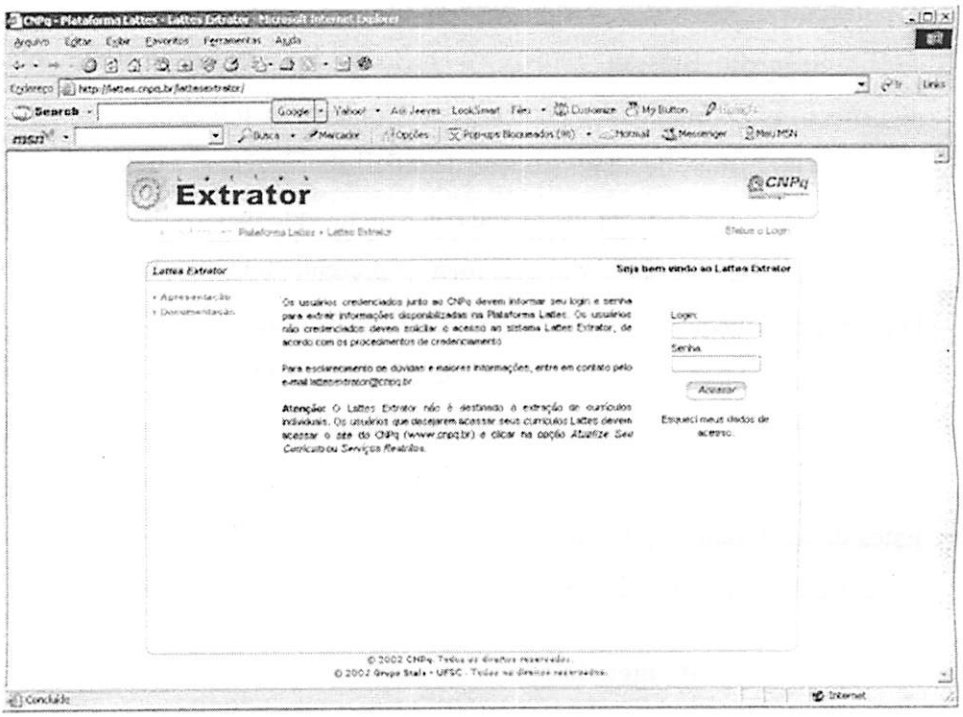


FIGURA 3.1 Interface disponível na internet do Lattes Extrator
Fonte: Grupo Stela, 2002b.

Os 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. Uma descrição da estrutura

dessas tabelas encontra-se no Apêndice A e o modelo Entidade-Relacionamento, que deu origem ao modelo relacional, encontra-se no Apêndice B. Entre as tabelas, algumas das principais e que merecem um destaque, podem ser citadas: dados gerais, com endereço profissional, formação acadêmica e atuações profissionais; produção bibliográfica, que inclui publicação de artigos e trabalho em eventos; produção técnica; outra produção e dados complementares, como participação em banca, orientações concluídas, entre outras.

Vale ressaltar que os dados coletados, em sua grande maioria, são do tipo nominal, o que torna mais difícil a tarefa de realizar análises estatísticas. Estes dados são mais complexos e considerados semi-estruturados. Segundo Abiteboul (1997), os dados semi-estruturados têm as seguintes principais características:

- definição *a posteriori*: os esquemas para dados semi-estruturados são usualmente definidos após a existência dos dados, com base em uma investigação de suas estruturas particulares e da análise de similaridades e diferenças. Isto não significa que sempre existe um esquema associado a um dado semi-estruturado, podendo não haver;
- estrutura irregular: coleções extensas de dados semanticamente similares estão organizadas de maneiras diferentes; algumas ocorrências podem possuir informações incompletas ou adicionais em relação a outras. Em suma, não existe um esquema padrão para esses dados. Os currículos extraídos da Plataforma Lattes enquadram-se nesta característica;
- estrutura implícita: muitas vezes, existe uma estrutura básica para os dados; porém, essa estrutura está implícita na forma como os dados são apresentados. É necessário realizar uma computação para obter essa estrutura;
- estrutura parcial: apenas parte dos dados disponíveis pode ter alguma estrutura, seja implícita ou explícita. Como conseqüência, um esquema

para estes dados nem sempre é completo do ponto de vista semântico e nem sempre todas as informações esperadas estão presentes;

- estrutura extensa: a ordem de magnitude de uma estrutura para estes dados é grande, uma vez que os mesmos são muito heterogêneos. Em outras palavras, os currículos podem ser preenchidos em diferentes formatos e a união destes dados pode produzir um esquema extenso;
- estrutura evolucionária: a estrutura dos dados modifica-se tão freqüentemente quanto os seus valores. Dados disponíveis na internet apresentam este comportamento, uma vez que existe o interesse em manter dados sempre atualizados;
- estrutura descritiva e não prescritiva: dada a natureza irregular e evolucionária dos dados semi-estruturados, as estruturas de representação implícitas ou explícitas normalmente se restringem a descrever o estado corrente de poucas ocorrências de dados similares. Um sinônimo para estrutura descritiva é estrutura indicativa, ou seja, nem sempre descreve exatamente, mas indica uma descrição.

As características de dados semi-estruturados diferem bastante das características de dados mantidos em bancos de dados tradicionais, como é o caso do Banco de Dados (BD) relacional criado neste trabalho. A Tabela 3.1 apresenta estas diferenças.

TABELA 3.1 Diferenças entre dados tradicionais e dados semi-estruturados.

Dados tradicionais	Dados semi-estruturados
Esquema predefinido	Nem sempre há um esquema predefinido
Estrutura regular	Estrutura irregular
Estrutura independente dos dados	Estrutura embutida no dado
Estrutura reduzida	Estrutura extensa
Estrutura fracamente evolutiva	Estrutura fortemente evolutiva
Estrutura prescritiva	Estrutura descritiva
Distinção entre estrutura e dado é clara	Distinção entre estrutura e dado não é clara

Fonte: Elaborada pela autora, com base em Abiteboul, 1997.

Um dado semi-estruturado possui um contexto que o envolve e o faz possuir características que se assemelham mais ao conceito de informação do que ao conceito de dado.

Bancos de dados tradicionais apresentam um esquema predefinido e uma estrutura homogênea para os tipos de dados. Já nos dados semi-estruturados, cada ocorrência de dado pode ser heterogênea. Dada essa heterogeneidade, em geral, a estrutura de um dado semi-estruturado está presente na própria descrição do dado, necessitando ser identificada e extraída. Estas tarefas são complexas, uma vez que a distinção entre esquema e dados nem sempre é clara, se forem comparadas ocorrências de dados semanticamente iguais. Foi necessária a criação de um BD tradicional, baseado nestes dados semi-estruturados.

Implementado o BD tradicional, este foi povoado com mais de 28 mil linhas de dados. Para o povoamento dos dados, cada currículo extraído preenchia diversas tabelas do BD, aquelas cujos dados das pessoas estavam cadastrados no currículo. Posteriormente, passou-se à etapa de filtragem, quando foram removidos dados inconsistentes, campos em branco, informações repetidas, entre outros. Após a filtragem, iniciou-se a etapa de mineração dos dados. Esta etapa foi subdividida em duas fases: a primeira, consistindo basicamente de cruzamento de consultas simples SQL[9]; a segunda, consistindo de funções e procedimentos que executam as técnicas mais específicas de *Data Mining*.

Todas estas tarefas foram desenvolvidas de forma específica para conhecer os dados inseridos e gerar relatórios na forma de gráficos, permitindo sua análise e interpretação, e resultando em conhecimento sobre a pesquisa científica na UFLA.

---

[9]*Structured Query Language* - Linguagem estruturada de consultas e manipulação em Banco de Dados (Elmasri &Navathe, 2002)

3.3 Desenvolvimento

Um grande obstáculo relativo à definição do problema do qual este trabalho trata, e que foi identificado no decorrer do seu desenvolvimento, é 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 outros.

Dentre as etapas pré-definidas da técnica de Descoberta de Conhecimento em Bancos de Dados (DCBC), apresentadas na Seção 2.2, foram realizadas:

1. seleção dos dados: por meio do Lattes Extrator, foram selecionados e extraídos, inicialmente, mais de mil documentos da Plataforma Lattes, que continham os registros de toda produção científica dos docentes, de alguns alunos, ex-alunos, mestrandos e doutorandos da UFLA, dentre outras pessoas. Em seguida, foram selecionados 575 currículos, dentre estes documentos, que continham dados específicos referentes às produções científica, tecnológica e bibliográfica dos mesmos, principalmente dos professores;

2. pré-processamento dos dados: realizada a partir da eliminação de incongruências e ou erros dos dados (filtragem). 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, o banco de dados resultante passou a conter 28.389 linhas. Nesta etapa do processo de DCBD não foi realizado o enriquecimento dos dados pelo fato de eles serem referentes a outras pessoas, extraídos dos documentos disponíveis na Plataforma Lattes, que já continham as informações necessárias à descoberta de conhecimento proposta;

4. transformação dos dados: foram feitos dois tipos de codificação de dados. O primeiro consistiu da transformação dos documentos obtidos no formato XML (dados semi-estruturados) em documentos SQL (BD relacional), contendo o código de inserção 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 Sistema Gerenciador de Bancos de Dados (SGBD) da Oracle. Esta etapa será mais detalhada posteriormente, ainda nesta seção;

5. *Data Mining*: a etapa consistiu da elaboração de algumas tarefas de *Data Mining*, pela implementação de técnicas específicas para este fim, realizando-se o cruzamento e a comparação de consultas e funções definidas na linguagem de programação PL/SQL, própria do SGBD Oracle;

6. interpretação: a interpretação dos resultados obtidos, que gera o conhecimento, é demonstrada a partir da criação de relatórios. O principal relatório desenvolvido é esta dissertação que contém, além de todo o referencial teórico acerca do tema, os resultados apresentados de diversas formas, desde gráficos resumidos até a descrição dos principais resultados, discutidos no Capítulo 4.

A Figura 3.2 foi adaptada da Figura 2.4, da Seção 2.2. Ela foi modificada para ilustrar passo a passo a identificação dessas etapas realizadas com as etapas pré-definidas do processo de DCBD:

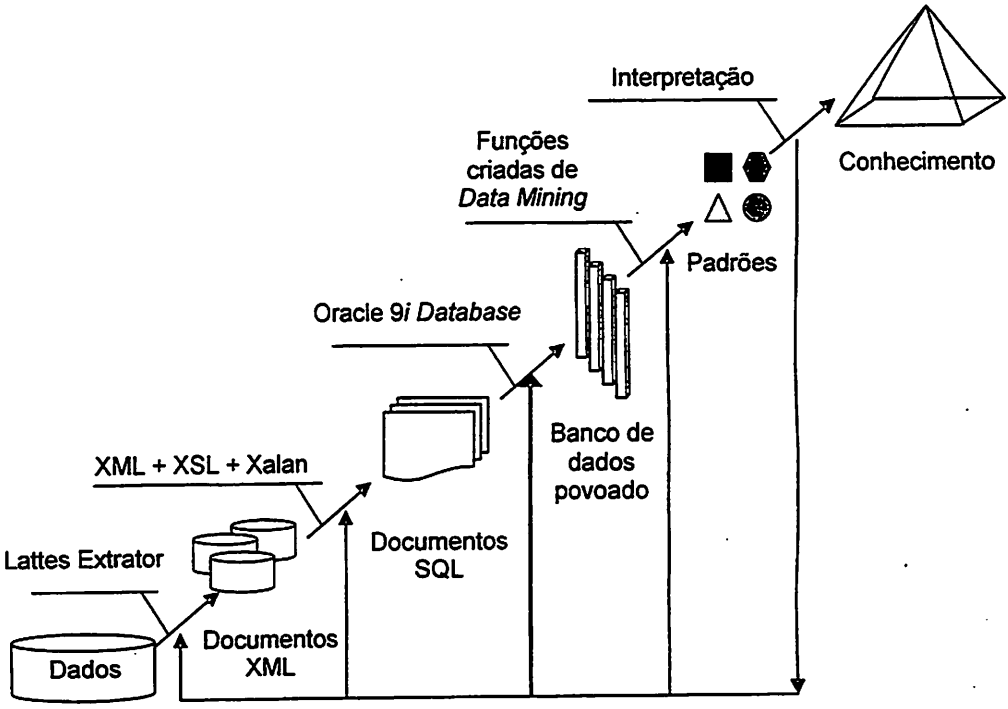


FIGURA 3.2 Execução das etapas do processo de DCBC neste trabalho
Fonte: Elaborado pela autora adaptada de Santos (2002).

A primeira etapa de seleção dos dados foi realizada a partir da ferramenta Lattes Extrator, que gerou como documentos a serem analisados, documentos em XML. Na etapa de pré-processamento, foi realizada uma codificação dos dados e, para tanto, foi necessário desenvolver um programa que extraísse as informações contidas nos documentos XML, que são dados semi-estruturados e as inserisse em um banco de dados tradicional. 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, versão Oracle 9*i* *DataBase*. Isso gerou os dados pré-processados em formato de documentos SQL.

Foi escolhido o SGBD da Oracle pela facilidade de uso, potência e uma boa relação custo-desempenho. O Oracle 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*.

Todos os programas foram desenvolvidos e utilizados em uma máquina com a seguinte configuração: processador Duron de velocidade de processamento de 1GHz, com disco rígido de 40GB e com 256 MB de memória RAM, que se encontra no Departamento de Ciência da Computação da UFLA.

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

3.3.1 XML - eXtensible Markup Language

A *eXtensibel Markup Language* (XML) é uma linguagem de marcação semelhante à HTML[10], porém, mais flexível no que se refere às marcações (*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 do próprio dado e um valor associado a ele. 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, mas não otimizados para tratamento por computadores (Ramalho, 2002).

Segundo Deitel (2003), a XML é uma linguagem de marcação que descreve dados de praticamente qualquer tipo, de forma estruturada

[10] HTML – *HiperText Markut Language*: linguagem de marcação e exibição de dados, principalmente utilizada para construção de páginas na internet.

hierarquicamente. Diferentemente da HTML, que possui um conjunto fixo de marcações, a XML permite descrever os dados de forma mais precisa, por meio da criação de novas marcas.

Contudo, a XML não está limitada a aplicações voltadas para a internet. 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.

Segundo Ramalho (2002), atualmente, diversas organizações de grande 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; organizações podem acessar dados de parceiros, de forma a diminuir os custos com a transmissão e compartilhamento de dados; *sites* da internet podem oferecer parte de seu conteúdo para ser inserido em outros *sites*, aumentando o número de visitas e etc.

Utilizou-se a linguagem XSL para identificar a estrutura dos documentos XML e depois, então, extrair destes documentos as informações necessárias para serem inseridas no banco de dados.

3.3.2 XSL - eXtensible Stylesheet Language

A *eXtensible Stylesheet Language* (XSL) é uma linguagem que permite transformar documentos XML em diversos outros formatos, como HTML, texto simples ou qualquer outro documento baseado em texto. A transformação é obtida associando-se padrões com gabaritos. Um padrão é confrontado com elementos de uma fonte de dados e um gabarito é gerado para criar parte do resultado. O resultado é separado da fonte; por isso, a estrutura do resultado pode ser completamente diferente da estrutura da fonte. Na construção do resultado, os elementos da fonte podem ser filtrados e reordenados, podendo-se ainda adicionar estrutura arbitrária (W3C, 1999).

Em outras palavras, a XSL é uma linguagem para a criação de folhas de estilo[11] para documentos XML. Por possuir essa funcionalidade, a linguagem XSL foi utilizada, juntamente com o programa Xalan, citado anteriormente, para criar uma folha de estilo para o currículo extraído da Plataforma Lattes.

A folha de estilo identificou cada *tag* do documento XML, extraindo seus valores e estruturando-os num código SQL que, depois, foi executado no SGBD da Oracle, para povoar o Banco de Dados.

O arquivo XSL deve conter todas as instruções necessárias para a conversão, como identificação das *tags*, dos atributos e dos valores destes. Neste trabalho, o papel do XSL foi apenas o de auxiliar, ou seja, comparar os dois documentos, fazer uma associação entre as estruturas e identificar os valores. O arquivo de saída é um documento no formato SQL.

A Figura 3.3 mostra uma parte do arquivo XML, extraída do currículo de "Olinda Nogueira Paes Cardoso", utilizado como exemplo.

```
<?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
Fonte: Grupo Stela, 2002b.

[11] Folha de Estilo ou CSS (*Cascadding Style Sheet*) é um recurso que define modelos de formatação de uma página de internet (Deitel, 2003).

O programa para a transformação de documentos XML em outros formatos (Xalan) trabalha associando o documento a uma folha de estilo específica desse XML e que encontra-se no formato XSL. Para cada currículo, foi automaticamente gerado um código XSL, criado para a extração das suas informações.

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 gerado para o exemplo apresentado anteriormente é como mostrado na Figura 3.4.

```
<?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>
```

FIGURA 3.4 Parte do código XSL utilizado na conversão
Fonte: Elaborado pela autora.

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 símbolo @, que deverá ser inserido no banco de dados. O arquivo de saída é automaticamente gerado, como ilustrado na Figura 3.5.

```
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, '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 código na linguagem SQL gerado a partir da conversão
Fonte: Elaborado pela autora.

Antes da inserção dos dados, foi necessária a modelagem e a criação do Banco de Dados. Para tanto, utilizou-se o SGBD Oracle, versão 9*i*, como citado anteriormente.

Depois da modelagem, criação e ajuste do Banco de Dados, foram executados, no SGBD Oracle 9*i*, 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.

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. Como exemplo de alguns campos que apareceram em branco em alguns documentos, tem-se: 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 etapa de mineração dos dados com o objetivo de conhecer melhor as relações existentes entre os dados dos currículos. Primeiramente, foram definidas algumas consultas mais simples que foram executadas sobre o banco de dados, no intuito de identificar padrões, associações e regras que constituiriam a implementação da ferramenta de *Data Mining* numa fase posterior a esta. Algumas dessas consultas seguem como exemplo:

- Quantas pessoas cadastradas possuem vínculo profissional com a UFLA?
- Que tipos de vínculos profissionais são permitidos na UFLA?
- Quais as áreas de maior atuação profissional?
- Quantas áreas do conhecimento diferentes foram cadastradas?
- Existem produções científicas de várias pessoas numa especialidade?
- Quais são as linhas de pesquisa existentes, hoje, na UFLA?
- A que órgãos estão ligadas às 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 maior publicação de artigos da UFLA, no Brasil ou no exterior?
- Em quantas linhas de pesquisa se encaixa cada atividade de uma pessoa?
- Quantas das atuações profissionais de cada pessoa são atividades de ensino?

- Qual o número máximo e o mínimo de atuações profissionais de uma 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 outras consultas foram realizadas utilizando-se uma interface de linha de comando do SGBD Oracle, chamada SQLPlus. Algumas dessas linhas de comando utilizadas para fazer as consultas estão ilustradas na Figura 3.6.

```
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) 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
Fonte: Elaborado pela autora.

Após esta fase inicial, foram feitas consultas mais aprofundadas, que geraram melhores resultados para o *Data Mining*. Para serem implementadas estas consultas no SGBD, foram criadas funções mais complexas. Dentre estas, pode-se citar como exemplo: Qual a implicação de uma atividade de direção para outras atuações, tais como atividades de pesquisa, trabalhos em eventos, serviços técnicos, número de orientações e de disciplinas ministradas? Existe uma relação entre o tempo de serviço de uma pessoa na UFLA e a quantidade de publicações realizadas? Existe uma relação entre o tipo de atuação desempenhada pelas pessoas (direção, pesquisa e ensino) e a quantidade de publicações destas pessoas? Existe alguma relação entre a publicação de um artigo no exterior e o fato do autor ter sido pós-graduado fora do Brasil?

Esses são alguns exemplos do vasto campo de conhecimento que pode ser descoberto pela utilização do sistema criado neste trabalho e pela atualização do processo de DCBD. Estas e outras consultas serão melhor detalhadas no Capítulo 4, juntamente com a análise dos resultados obtidos. Outras descobertas e novas possibilidades de execução do *Data Mining* compreendem sugestões para trabalhos futuros.

4 O ESTUDO EMPÍRICO: RESULTADOS E ANÁLISE

4.1 Gestão de ciência, tecnologia e inovação na UFLA

Esta seção tem como principal objetivo situar as condições da pesquisa na Universidade Federal de Lavras. Para tanto, foi feita uma descrição da UFLA com um resumo das principais ações, tomadas no período de 2000 a 2004, com relação ao desenvolvimento de CT&I nesta universidade.

Esta seção está baseada em informações colhidas na Pró-Reitoria de Pesquisa (PRP) da UFLA, por meio da pesquisa a documentos produzidos pela própria PRP (PRP, 2004).

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, oferece 10 cursos de graduação e 28 cursos de pós-graduação presenciais (destes, 14 de mestrado, 12 de doutorado e 2 de especialização).

Diretamente ligados 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 1960, quando houve grande preocupação com a expansão e qualificação do quadro de docentes, o que possibilitou a criação dos cursos de pós-graduação a partir de 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 com aproximadamente 65 grupos 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 nas agências de fomento para as atividades de 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 essa atividade e aumentar sua aplicabilidade e inserção na sociedade.

As políticas e ações da pesquisa na UFLA evoluíram substancialmente nas últimas décadas, devido, principalmente, a fatores como 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 a substituição do quadro de docente que, de 1996 a 1999, foi superior a 53%; a consolidação dos programas de pós-graduação e a ampliação e melhoria de infra-estrutura.

Desenvolver pesquisa é a grande motivação e incentivo dos docentes, devido à valorização pessoal e profissional; à 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; à progressão funcional da carreira do docente; ao incentivo financeiro; a

possibilidades de assessoria/consultoria, como tarefas de extensão; ao reforço financeiro para o sistema, advindo de auxílios externos e à facilitação de inserção na comunidade, que é missão social da universidade.

A agenda de trabalho da UFLA neste período está direcionada à organização interna e a ações extra-campus, visando ampliar a dimensão da pesquisa, tanto em termos de quantidade quanto de qualidade e promover maior inserção da mesma na sociedade. A seguir são 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 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 mesma;

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[12], visando ampliar as oportunidades de parcerias externas, no âmbito da C&T (projetos, assessorias, consultorias e serviços);

[12] Fundação de Apoio ao Ensino, Pesquisa e Extensão (http://www.faepe.org.br/).

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 PRP/UFLA criou Comissões de Assessoramento, cada uma com sua competência bem definida. 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; 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 (NITs) 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 à 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 um grande aumento no número de grupos de pesquisa, 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 à composição e definição de linhas de pesquisa, para adequar-se aos critérios do órgão gestor (CNPq) e à 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 do maior número possível de grupos de pesquisa, na relevância econômica e social da atividade, especialmente nas áreas de maior inserção, levando-se em consideração a tradição, a capacitação e a localização geográfica da UFLA, na necessidade de que a universidade se envolva mais com pesquisas direcionadas à solução de problemas evidentes e avanços sócio-econômicos e ambientais e na demanda real por 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 suas competências, em linhas de pesquisa e interesses específicos que se aglomeram, por temáticas comuns, em subprojetos e projetos independentes ou integrados, até mesmo extra-institucionais. Cada programa institucional poderá envolver ou resultar de vários projetos ou subprojetos 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 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 busquem-se alguns ajustes de foco 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 se 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 do impacto destes para C&T, nas últimas décadas do século XX e difundir e ampliar a sua participação 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.

4.2 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 (Grupo Stella, 2002b).

A Plataforma Lattes é um conjunto de sistemas de informação, bases de dados e portais da internet, concebida para integrar os sistemas de informação das agências federais, racionalizando o processo de gestão de C&T. Lançada em 16 de agosto de 1999, proporcionou um aumento significativo do número de currículos enviados ao CNPq, que chegou a mais de 100 por dia. Segundo dados do Grupo Stela (2002b), a Plataforma Lattes possui aproximadamente 480.000 currículos cadastrados[13].

[13] Informação extraída da página disponível em http://lattes.cnpq.br, em 28/12/2004.

Criado pela Lei nº 1.310 de 15 de janeiro de 1951, o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) é uma fundação, vinculada ao Ministério da Ciência e Tecnologia (MCT), para o apoio à pesquisa brasileira. Contribuindo diretamente para a formação de pesquisadores (mestres, doutores e especialistas em várias áreas de conhecimento), o CNPq é, desde sua criação até hoje, uma das maiores e mais sólidas estruturas públicas de apoio à Ciência, Tecnologia e Inovação (CT&I) dos países em desenvolvimento.

Os investimentos feitos pelo CNPq são direcionados para a formação e absorção de recursos humanos e financiamento de projetos de pesquisa que contribuem para o aumento da produção de conhecimento e geração de novas oportunidades de crescimento para o país.

A função de fomento constitui-se na principal ação desenvolvida pelo CNPq, com vistas à promoção do desenvolvimento científico e tecnológico do país. Como linha de trabalho mais tradicional e identificadora da missão do órgão, o fomento é dirigido essencialmente para a formação de recursos humanos e para o apoio à realização de pesquisas.

No primeiro caso, a ação desenvolvida destina-se a gerar uma capacitação científica e tecnológica nacional, pela formação de pesquisadores altamente qualificados. O apoio à pesquisa expressa, por sua vez, o cumprimento de uma responsabilidade do Estado em promover e estimular a produção de conhecimentos necessários ao desenvolvimento econômico e social, à afirmação da identidade cultural e ao aproveitamento racional dos recursos naturais do país.

Para a implementação dessas ações, o CNPq opera um conjunto de instrumentos – bolsas no país, bolsas no exterior e fomento à pesquisa (ou auxílio à pesquisa) – e de suas diversas modalidades.

Com o objetivo principal de dar transparência às ações relacionadas à sua função primordial, o CNPq disponibiliza informações sobre o histórico de

bolsistas e de pesquisadores beneficiados nos dez últimos anos, por área do conhecimento, instituição, unidade da federação e modalidade da bolsa ou auxílio.

A Plataforma Lattes constitui um importante passo para a integração dos sistemas de informação das principais agências de fomento nacionais, antiga demanda da comunidade científica e tecnológica. O Sistema CV-Lattes, em suas versões *on-line* e *off-line*, é o componente da Plataforma Lattes desenvolvido para o CNPq e utilizado por MCT, FINEP, CAPES e por todos os atores institucionais, bem como pela comunidade científica brasileira como sistema de informação curricular.

Fazem uso desse sistema: pesquisadores, estudantes, gestores, profissionais e demais atores do Sistema Nacional de Ciência, Tecnologia e Inovação. No CNPq, suas informações são aplicadas:

1. na avaliação da competência de candidatos à obtenção de bolsas e auxílios;
2. na seleção de consultores, de membros de comitês e de grupos assessores;
3. no subsídio à avaliação da pesquisa e da pós-graduação brasileiras.

Para que esses objetivos possam ser alcançados de forma plena, o CNPq decidiu que, a partir do ano de 2002, todos os bolsistas de pesquisa, de mestrado, de doutorado e de iniciação científica, orientadores credenciados e outros clientes do Conselho teriam de ter seu currículo cadastrado na Plataforma Lattes do CNPq. A inexistência do currículo impediria pagamentos e renovações. O currículo também seria obrigatório para todos os pesquisadores e estudantes que participam do Diretório de Grupos de Pesquisa no Brasil. Apesar disso, esta obrigatoriedade não se estabeleceu até os dias atuais, mas, a qualquer momento, os interessados (bolsistas, pesquisadores e estudantes) podem criar ou atualizar seus currículos e enviá-los ao CNPq.

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.

4.2.1 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, por meio 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, disponibilizado pelo CNPq (Grupo Stela, 2002a).

Além disso, as instituições podem desenvolver suas próprias rotinas para importação dos dados curriculares para suas bases corporativas, uma vez que eles 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.

Neste ano de 2004, 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 arquivos em XML dos currículos recebidos após o encerramento do Censo 2002, esses estarão sendo disponibilizados para extração. O acesso ao Lattes Extrator depende do credenciamento prévio, pelo CNPq, da instituição interessada.

4.3 Resultados e discussões

O pressuposto inicial de que há uma grande quantidade de informação e conhecimento "escondidos" nos registros da pesquisa científica da UFLA é bastante válido, uma vez que a riqueza de informações obtidas a partir das respostas alcançadas com as consultas poderia ser mais aproveitada pelos órgãos de direção da universidade envolvidos na pesquisa científica.

Verificou-se que os dados presentes nos currículos extraídos da Plataforma não estavam atualizados, o que representou uma limitação para este trabalho. Até o presente momento, as informações disponíveis no *site* oficial do CNPq são de que a versão do Lattes Extrator que está disponível extrai apenas currículos atualizados até julho de 2002. De acordo com o *site*, está sendo

desenvolvida uma nova versão que permitirá a extração de currículos mais atualizados, logo que estiver disponível (Grupo Stela, 2002b). Sendo assim, apesar desta limitação, uma vez disponibilizados novos dados, o mesmo trabalho poderá ser realizado, apenas executando-se as funções já criadas para gerar conhecimeÉto mais atualizado.

4.3.1 Resultados superficiais

Os primeiros resultados obtidos neste trabalho foram conseguidos por meio de consultas simples realizadas diretamente sobre o banco de dados dos currículos, utilizando para tanto a linguagem de programação PL/SQL, do SGBD Oracle.

Um dos grandes problemas encontrados para realizar a análise dos dados é a falta de padronização dos valores cadastrados. Por exemplo, existem 72 cargos diferentes, 46 órgãos diferentes e 172 unidades distintas cadastrados nos currículos de pessoas ligadas à UFLA. Muitos destes dados, na realidade, representam um mesmo objeto, tal como o Departamento de Administração e Economia que pode, ao mesmo tempo, ser cadastrado como um órgão ou como uma unidade. E mais, este mesmo departamento poderia ser novamente cadastrado pela sigla DAE. Todas estas diferentes formas de cadastrar este objeto deveriam ser representados de forma única. Ocorrem também casos em que um mesmo objeto é cadastrado de forma redundante em tabelas diferentes, como é o caso de órgãos e unidades. No Apêndice C são apresentados dados que ilustram essa falta de padronização.

Outro problema refere-se ao próprio formato do currículo Lattes, que não deixa claro qual é a função de cada pessoa ligada à UFLA. Por exemplo, os dados referentes ao vínculo profissional das pessoas podem ser de seis tipos: celetista, colaborador, livre, outro, professor_visitante e servidor_publico. Observa-se que não há o vínculo definido como "professor", o que torna difícil a tarefa de afirmar com segurança quais são os professores da UFLA, pois existem

professores cadastrados como servidor_publico, livre ou outro. Sendo assim, considerou-se que uma pessoa é professor na UFLA quando a mesma possui atividades de ensino cadastradas em cursos oferecidos por esta instituição. Porém, não se pode afirmar com exatidão quem são as pessoas que não são professores na UFLA, pois podem existir casos de professores que não cadastraram suas atividades de ensino em seus currículos.

Além dessas limitações, outro fato que prejudicou a análise dos resultados gerados é que poucas pessoas atualizam seus currículos Lattes periodicamente e, quando o fazem, a maioria o faz de forma parcial.

Um resultado crítico que advém desse fato é que, dos 575 currículos inseridos no Banco de Dados, mais de 90% não contêm atividades cadastradas. As atividades podem ser de ensino, pesquisa, direção e extensão, além de serviços técnicos e treinamentos ministrados, que ocorreram ao longo dos anos, ou seja, uma só pessoa pode possuir, por exemplo, diversas atividades de direção cadastradas ao longo de toda a sua carreira. Os menos de 10% das pessoas, exatamente 55 pessoas, que incluíram suas atuações profissionais em seus currículos, têm entre 2 a 61 atuações, demonstrando uma variedade muito grande de número de atividades, chegando ao número total de 792 atuações distintas. O gráfico da Figura 4.1 demonstra essa variedade.

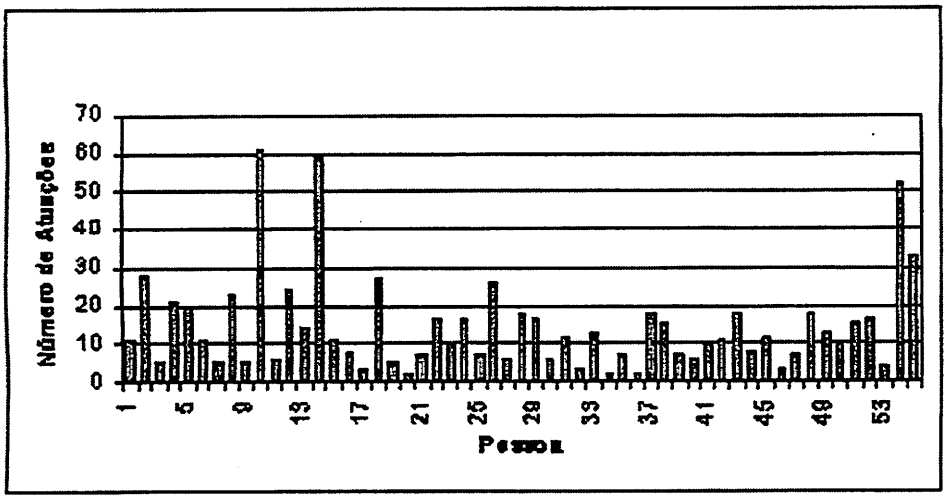


FIGURA 4.1 Número de atuações profissionais por pessoa
Fonte: Elaborado pela autora.

O que se pôde observar é que apenas 39 pessoas, aproximadamente 6% do total de currículos cadastrados no banco de dados, realizaram entre 1 a 16 atividades de ensino, de um total de 119 atividades cadastradas, como mostra o gráfico da Figura 4.2. Uma observação importante é que essas atividades de ensino erroneamente incluíam atividades de direção como, por exemplo, a gerência de organizações.

Entre as atividades de direção podem ser citadas: coordenação de curso, chefia e subchefia de departamento, coordenação de laboratório, etc. Apenas 3% das pessoas (19 pessoas) cadastraram alguma atividade de direção, sendo que, destas, de um total de 82 atividades cadastradas, há uma variação de 1 a 19 atividades por pessoa, cuja distribuição está ilustrada pelo gráfico da Figura 4.3.

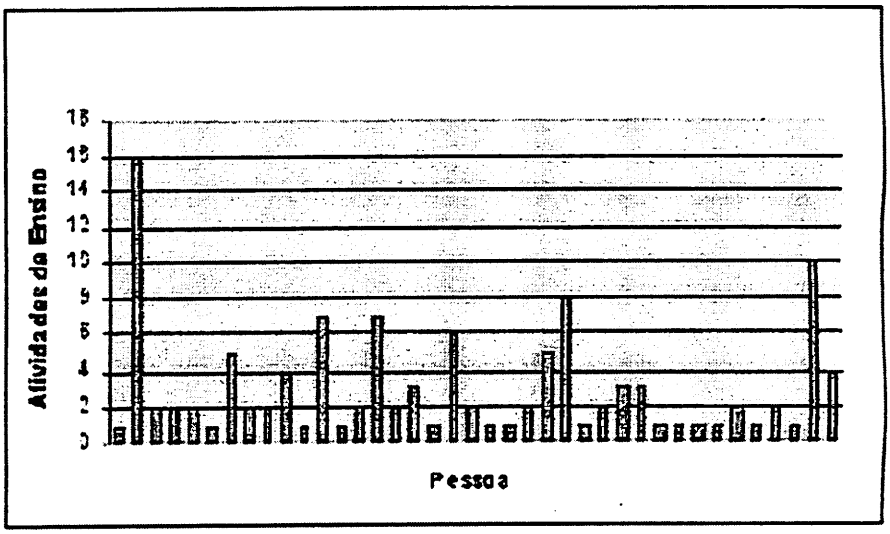


FIGURA 4.2 Número de atividades de ensino por pessoa
Fonte: Elaborado pela autora.

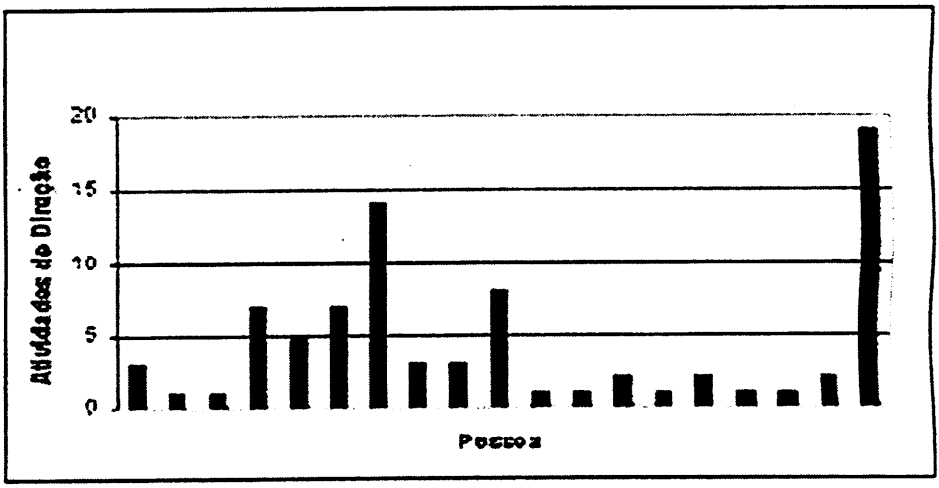


FIGURA 4.3 Número de atividades de direção por pessoa
Fonte: Elaborado pela autora.

No que se refere às atividades de pesquisa, 7% das pessoas (43 pessoas) realizaram entre 1 a 6 atividades de pesquisa, de um total de 86, como mostra o

gráfico da Figura 4.4. Cada uma delas compreendia, na maioria das vezes, de uma a três, mas algumas possuíam até 9 linhas de pesquisa.

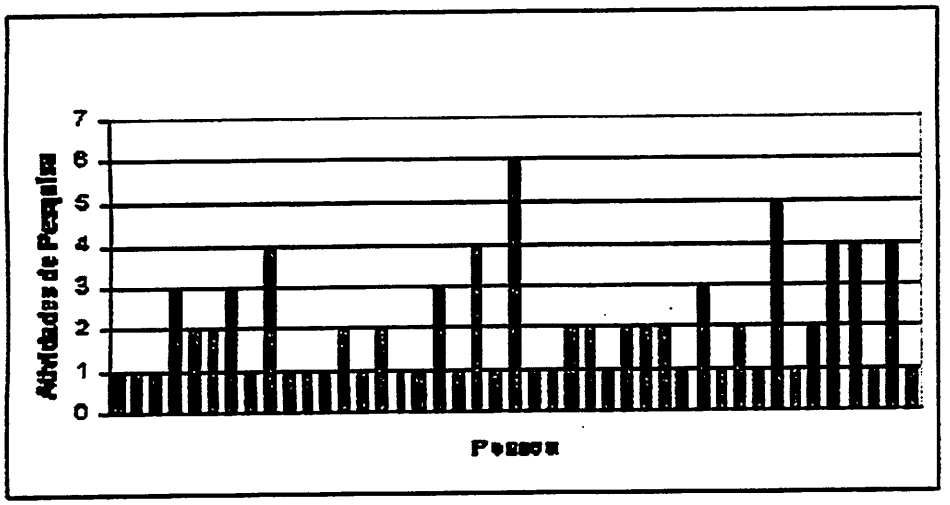


FIGURA 4.4 Número de atividades de pesquisa por pessoa
Fonte: Elaborado pela autora.

Além das atividades de pesquisa, 2,6% das pessoas cadastraram e realizaram entre 1 a 6 atividades de extensão, de um total de 22 atividades desse tipo, como mostra o gráfico da Figura 4.5. Neste exemplo, pode-se observar claramente um caso de *outlier*, ou seja, uma pessoa que foge totalmente do padrão e da média dos demais, contendo um número muito superior de atividades de extensão cadastradas.

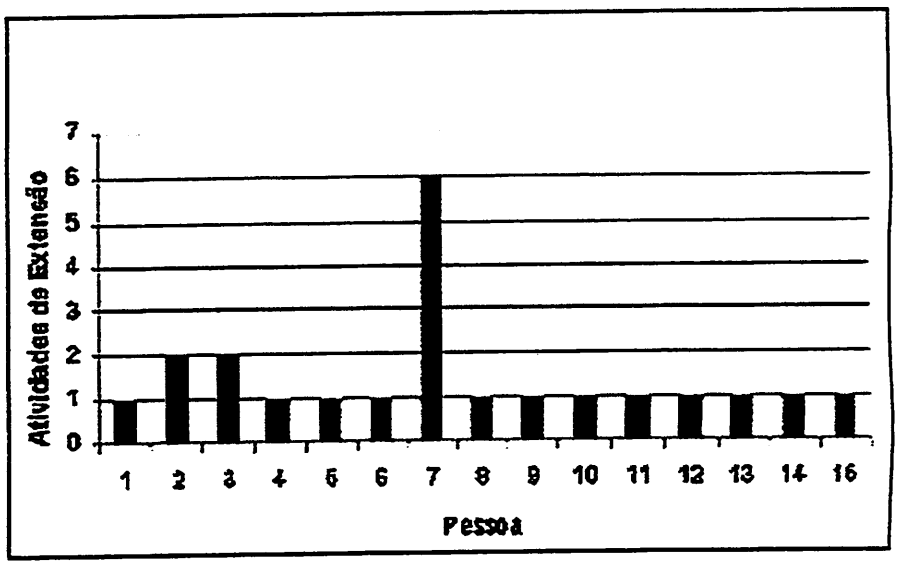


FIGURA 4.5 Número de atividades de extensão por pessoa
Fonte: Elaborado pela autora.

Além das atividades de ensino, direção, pesquisa e extensão já apresentadas, 2,4% das pessoas realizaram entre 1 a 32 serviços técnicos e menos de 2% ministraram algum treinamento, conforme distribuições ilustradas pelos gráficos das Figuras 4.6 e 4.7, respectivamente. Entre as atividades de serviço técnico, encontravam-se, por exemplo: atividades de assistência técnica, consultoria, levantamento planifotogramétrico de fazendas, dentre outros. Dos treinamentos ministrados, podem-se citar: cursos de capacitação, de reciclagem e de aperfeiçoamento, dentre outros exemplos.

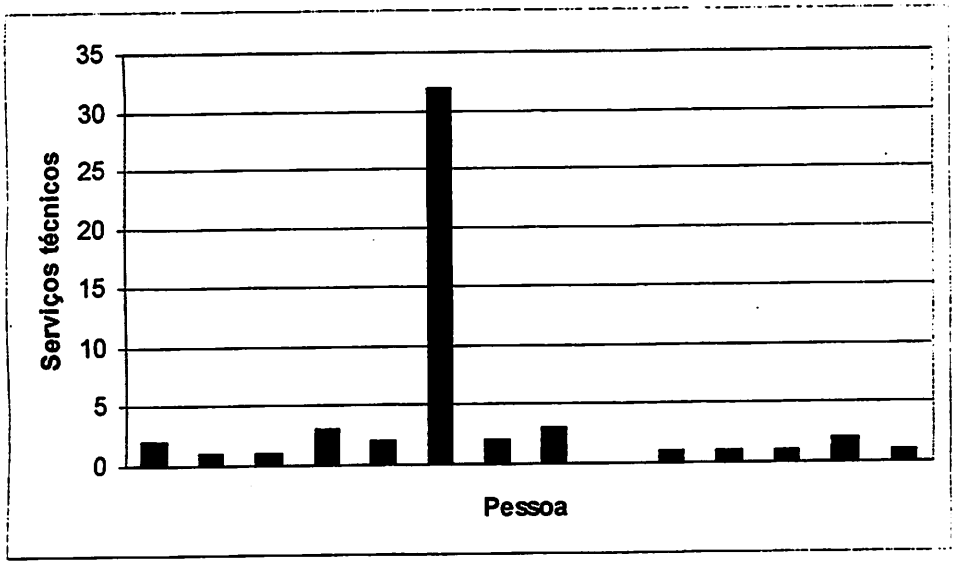


FIGURA 4.6 Número de serviços técnicos por pessoa
Fonte: Elaborado pela autora.

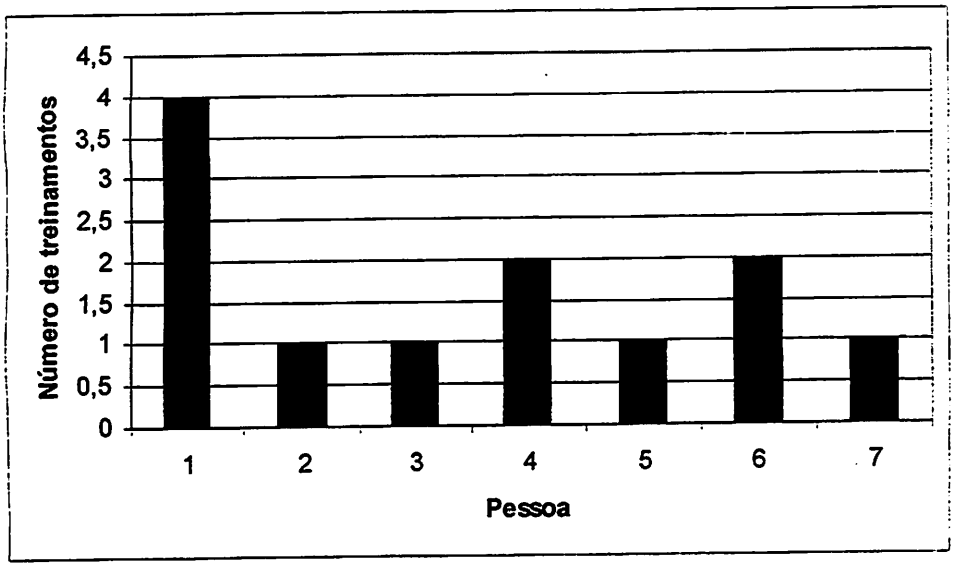


FIGURA 4.7 Número de treinamentos ministrados por pessoa
Fonte: Elaborado pela autora.

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.

Destes primeiros resultados, apenas observando-se os números de atividades cadastradas pelas pessoas, é interessante analisar que as mesmas, ao preencherem seus currículos na Plataforma Lattes, dão maior prioridade às atividades de ensino e pesquisa do que às demais.

Por outro lado, analisando-se as produções bibliográficas, observou-se que foram publicados 573 artigos de 1968 até o princípio de 2004, tendo a maior parte deles sido publicada em 2001. Vale ressaltar que como o banco de dados é oficialmente atualizado até julho de 2002, estranha-se o fato de haver publicações cadastradas até princípio de 2004. 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 4.8.

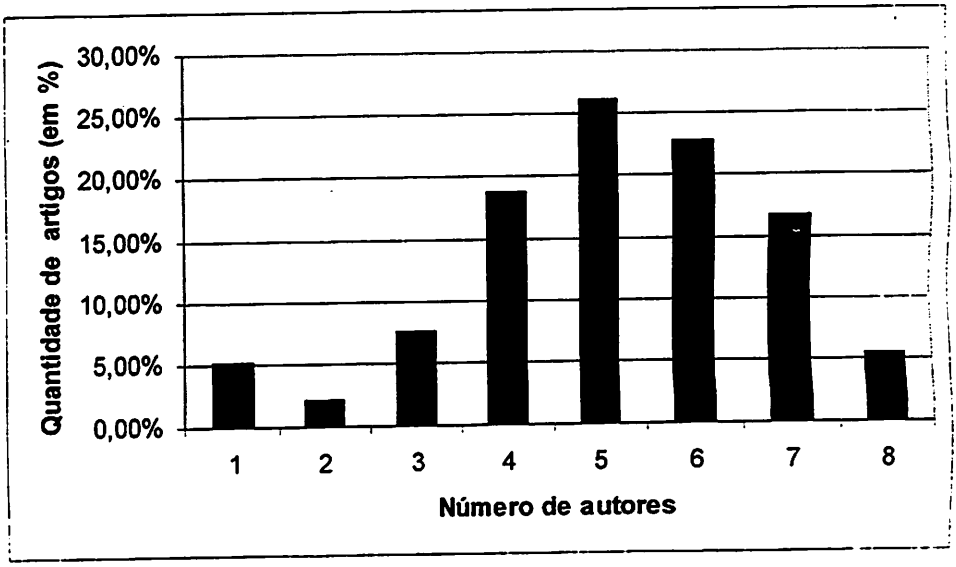


FIGURA 4.8 Número de autores por artigo
Fonte: Elaborado pela autora.

De acordo com o *site* oficial do CNPq, as grandes áreas do conhecimento são: Ciências Agrárias, Ciências Biológicas, Ciências da Saúde, Ciências Exatas, Ciências Humanas, Ciências Sociais Aplicadas, Engenharias e Lingüística. Todos os cursos de graduação, pós-graduação e especialização existentes hoje são pré-definidos pelo CNPq dentro dessas áreas (CNPq, 2004). As grandes áreas podem agregar vários cursos, tais como:

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

A Figura 4.9 mostra um gráfico com a distribuição dos artigos por área do conhecimento. Dos 573 artigos publicados, 77% pertencem à área de Ciências Agrárias, 13% à de Ciências Biológicas, 2,3% à de Ciências da Saúde, à de 5,4% Ciências Exatas, 0,3% Ciências Humanas, 1,7% Ciências Sociais Aplicadas e 0,3% Engenharias. Estes foram alguns dos primeiros resultados

obtidos com a aplicação do processo de descoberta de conhecimento em banco de dados.

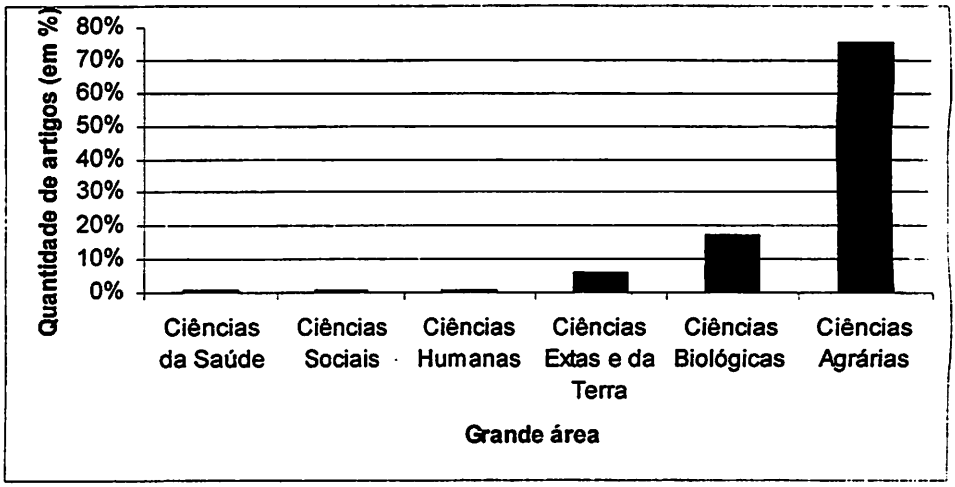


FIGURA 4.9 Número de artigos publicados por grande área do conhecimento
Fonte: Elaborado pela autora.

4.3.2 Resultados aprofundados

Com a utilização das técnicas de *Data Mining*, foram criadas funções específicas para descobrir padrões de comportamento mais relevantes nos dados disponíveis. Estes resultados são enquadrados nas categorias de conhecimento que podem ser geradas pela técnica de *Data Mining*, que estão na Seção 2.3.2.

4.3.2.1 Análises de regras de associação

Algumas funções de *Data Mining* foram criadas para analisar regras de associações entre elementos dos currículos existentes.

Um primeiro exemplo, ilustrado na Figura 4.10, mostra a associação entre a quantidade de publicações realizadas por pessoas que trabalham na UFLA e as que não trabalham. Esta função envolveu um total de onze tabelas do banco de dados, das quais sete são tabelas relacionadas às atuações e quatro são relacionadas às diversas formas de publicações. No total, foram obtidas 1.977

publicações; destas, 55% foram publicadas por pessoas que não estavam atuando na UFLA na época da publicação e 45% por pessoas que atuavam na UFLA na época da publicação.

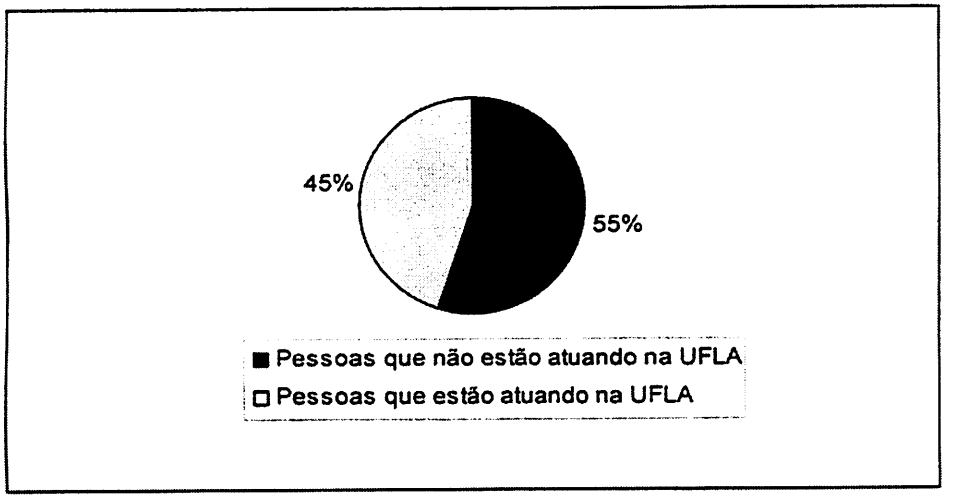


FIGURA 4.10 Associação entre o trabalho na UFLA e a quantidade de publicações
Fonte: Elaborado pela autora.

Vale analisar neste exemplo que uma pessoa, ao receber afastamento total para treinamento, fazer pós-graduação por exemplo, não está atuando na UFLA durante o período deste afastamento. Este fato poderia explicar o resultado encontrado, uma vez que no mestrado e ou doutorado, realiza-se mais pesquisa e publica-se mais. Isto também poderia refletir o fato de que, ao estar atuando na UFLA em atividades de ensino e direção, as pessoas podem ficar com a sua carga horária sobrecarregada e, conseqüentemente, acabem por realizar um número menor de pesquisas e publicações.

Outro exemplo, ilustrado na Figura 4.11, explora um pouco mais os resultados obtidos no exemplo anterior. Refere-se aos 55% das pessoas que não estavam atuando na UFLA, associados à quantidade de suas publicações nesse período de ausência. Esta função envolveu um total de sete tabelas do banco de dados, sendo três delas relacionadas às atuações e quatro relacionadas às

diversas formas de publicações. No total foram realizadas 1.062 publicações realizadas por pessoas que não estavam atuando na UFLA no momento da publicação.

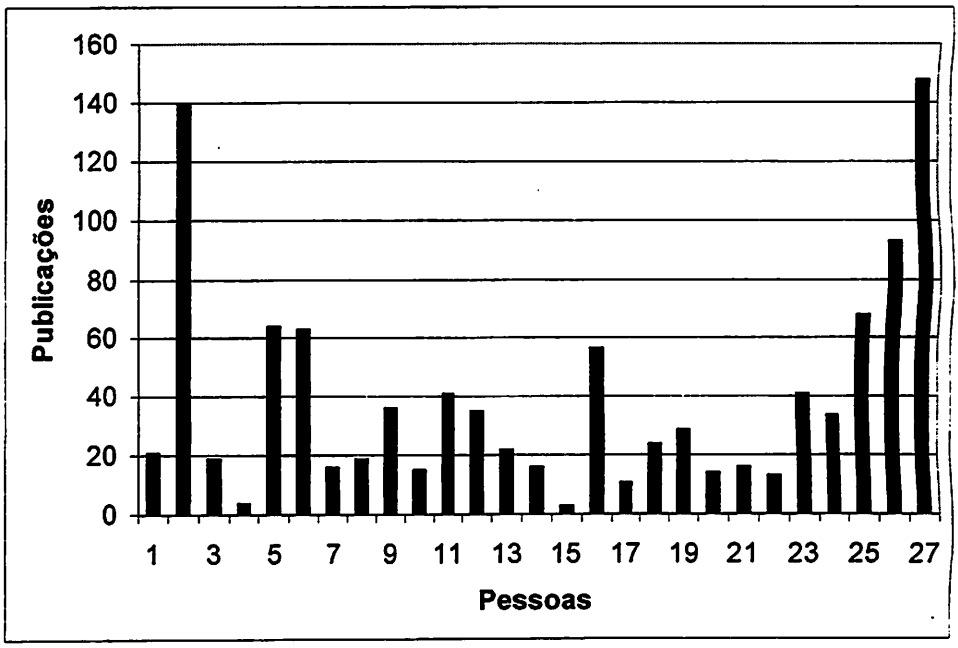


FIGURA 4.11 Quantidade de publicações de pessoas que não estavam atuando na UFLA no momento da publicação
Fonte: Elaborado pela autora.

Mais um exemplo de regra de associação mostra a relação entre todas as publicações cadastradas e o tempo de serviço de seus autores na UFLA, ilustrado na Figura 4.12. Esta função envolveu um total de onze tabelas do banco de dados, sendo sete delas relacionadas às atuações e quatro tabelas relacionadas às diversas formas de publicações. No total, foram obtidas 915 publicações relacionadas ao tempo de serviço de seus autores com a UFLA.

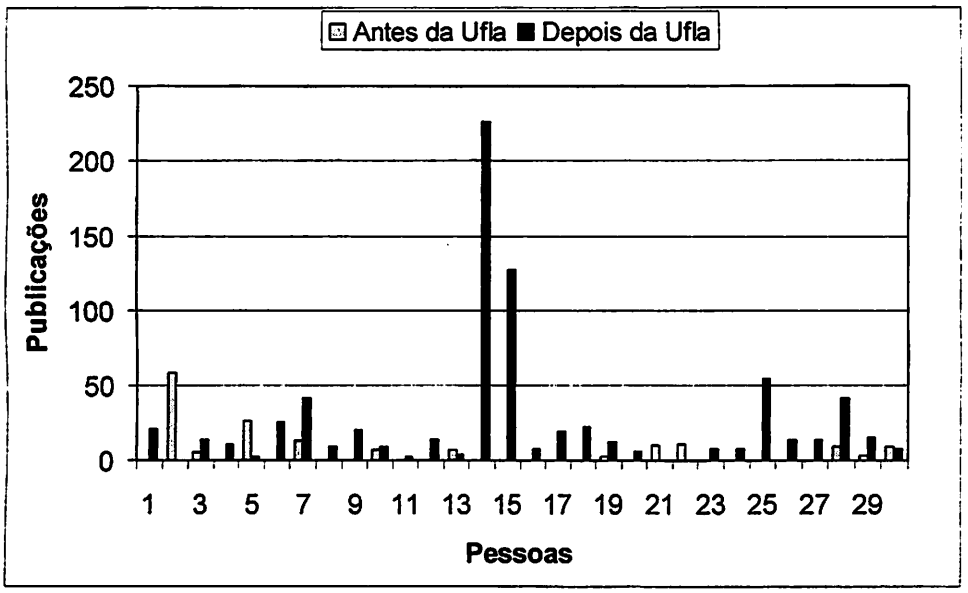


FIGURA 4.12 Relação entre publicações e tempo de serviço na UFLA
Fonte: Elaborado pela autora.

Analisando-se o exemplo ilustrado na Figura 4.12, percebe-se que a maioria das publicações feitas por pessoas que atuam na UFLA foi realizada depois que elas começaram a trabalhar na universidade.

Os dois próximos exemplos de regras de associação buscam mostrar a relação existente entre o fato das pessoas terem realizado pós-graduação no exterior ou no Brasil e o fato destas pessoas terem publicado no exterior.

A relação entre o local onde foi realizada a pós-graduação e o número de publicações no exterior é ilustrada na Figura 4.13. Esta função envolveu duas tabelas relacionadas à pós-graduação e quatro relacionadas aos tipos de publicações. No total, foram observadas 74 publicações realizadas no exterior por 42 pessoas, tendo a maioria destas publicações sido escrita por pessoas que fizeram pós-graduação no Brasil.

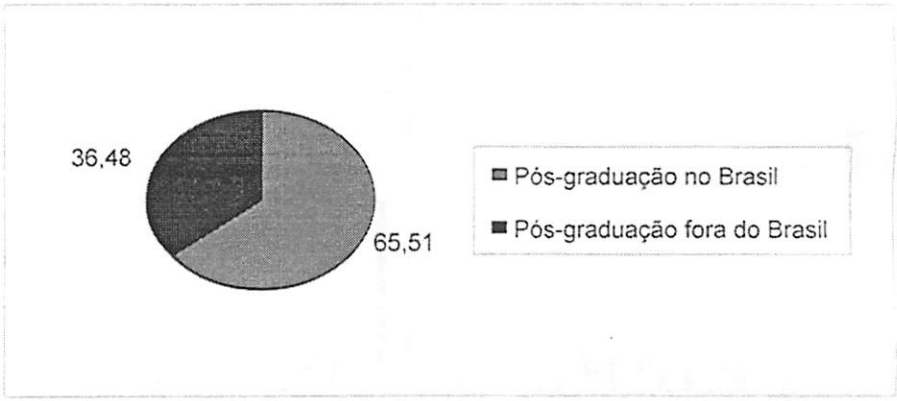


FIGURA 4.13 Relação entre o local da pós-graduação e o número de publicações no exterior
Fonte: Elaborado pela autora.

Este resultado deve-se ao fato de que, neste banco de dados, o número de pessoas que cursaram pós-graduação no Brasil (34 pessoas) é muito maior do que o das que cursaram no exterior (8 pessoas). Sendo assim, é natural que o número de publicações no exterior seja maior para este grupo de 34 pessoas do que para o outro. Porém, este resultado está ligado a outra medida que trata da média de publicações no exterior por cada pessoa. A Figura 4.14 mostra que a média de publicações no exterior de pessoas que cursaram a pós-graduação fora do Brasil é maior numa razão de 2,71 com relação às pessoas que cursaram pós-graduação no Brasil. A função que chegou a este resultado envolveu um total de seis tabelas, sendo duas relacionadas à pós-graduação e quatro relacionadas às publicações.

Esta relação indica que quem faz pós-graduação no exterior tende a ter maior visibilidade fora do Brasil, em termos de publicações, do que quem faz pós-graduação no Brasil.

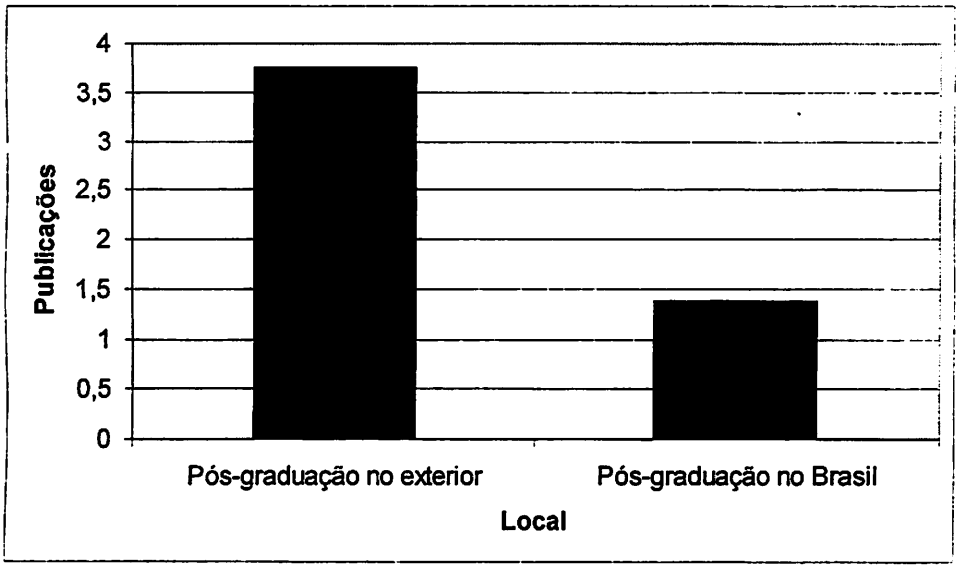


FIGURA 4.14 Média de publicações no exterior por pessoas e local de pós-graduação
Fonte: Elaborado pela autora.

4.3.2.2 Análises de regras de associação e *outliers*

No resultado apresentado na Figura 4.15, aparecem tanto a análise de regra de associação quanto uma análise de *outlier*. Pela Figura 4.15 percebe-se a presença de três pessoas com um número muito superior de linhas de pesquisa para suas atividades de pesquisa, podendo ser considerados *outliers*. Esta função envolveu tabelas relacionadas às linhas de pesquisa, grande área, área e subárea, e tabelas relacionadas às atividades de pesquisa desempenhadas. No total, foram obtidas 84 pesquisas e 186 linhas de pesquisa.

Vale esclarecer, para este banco de dados, a distinção que existe entre os termos "linha de pesquisa" e "atividade de pesquisa". No currículo Lattes, cada atividade de pesquisa na qual uma pessoa está envolvida durante um certo período (por exemplo, qualquer projeto de pesquisa envolvendo um grupo de pessoas ou isolado) pode estar ligada a uma ou mais linhas de pesquisa. As

linhas de pesquisa para cada atividade são definidas pelas pessoas ao preencherem seu currículo.

O mesmo fato ocorre com as grandes áreas, áreas e subáreas ligadas às atividades de pesquisa. Uma atividade de pesquisa deve possuir uma grande área e pode possuir uma ou mais áreas e subáreas associadas a ela. Uma pessoa não pode criar uma nova grande área e incluí-la em seu currículo. Porém, as áreas e subáreas não são pré-definidas, ou seja, uma pessoa pode criar uma nova área ou subárea para enquadrar sua atividade de pesquisa, Como estes campos são abertos no banco de dados, a tarefa de comparar estes dados é bastante complexa. O Apêndice D apresenta as áreas e subáreas cadastradas na Plataforma Lattes das pessoas ligadas à UFLA.

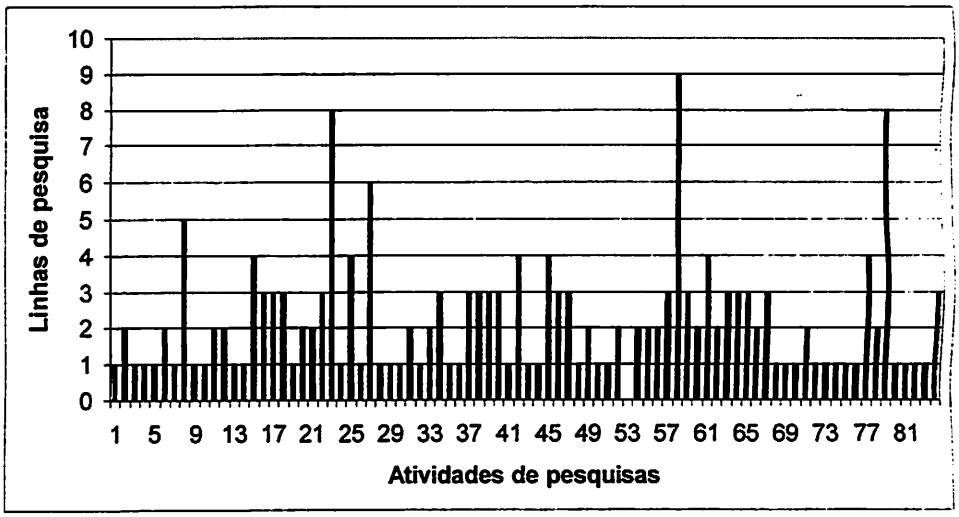


FIGURA 4.15 Associação entre as linhas de pesquisa e as atividades de pesquisa
Fonte: Elaborado pela autora.

4.3.2.3 Análises de regras de associação e de padrão seqüencial

A Figura 4.16 é resultado tanto da análise de regra de associação quanto da análise de padrão seqüencial. O objetivo da consulta era avaliar se havia uma

relação entre o tempo de conclusão do mestrado e o tempo de início do doutorado. Pela imagem percebe-se um padrão de comportamento, pois a maioria das pessoas leva entre 0 a 3 anos de intervalo entre estes dois tipos de pós-graduação. Nesta mesma consulta pôde-se observar a presença de *outliers* como pessoas que levaram mais de 20 anos entre o mestrado e o doutorado. Esta função envolveu a tabela contendo dados gerais das pessoas e duas tabelas sobre pós-graduação. No total, este resultado envolve 483 pessoas do banco de dados que cursaram mestrado e doutorado.

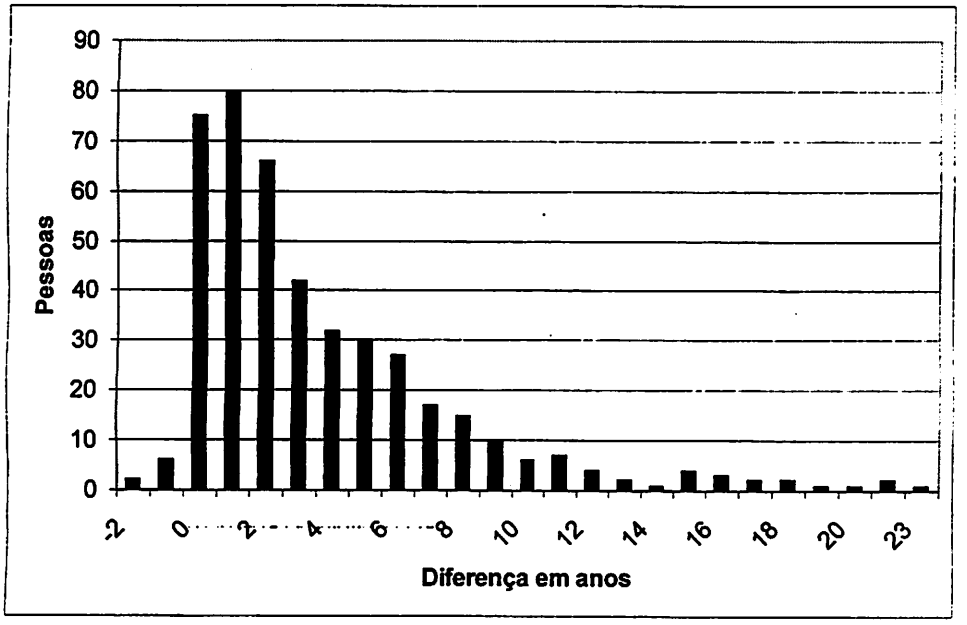


FIGURA 4.16 Associação temporal entre mestrado e doutorado
Fonte: Elaborado pela autora.

4.3.2.4 Análises de padrões seqüenciais

Os exemplos a seguir mostram padrões de comportamento seqüencial dos dados com relação ao tempo. A Figura 4.17 apresenta o resultado de uma consulta para avaliar se há uma entre o tempo de cadastramento do currículo e o tempo de vínculo profissional com a UFLA. Pelo gráfico, percebe-se um padrão

de comportamento, pois a grande maioria das pessoas cadastrou seu vínculo profissional com a UFLA a partir dos anos 1990. Nesta consulta, a função elaborada envolveu as tabelas de dados gerais das pessoas, as tabelas de atuações e a tabela de vínculo profissional, resultando um total de 82 ocorrências.

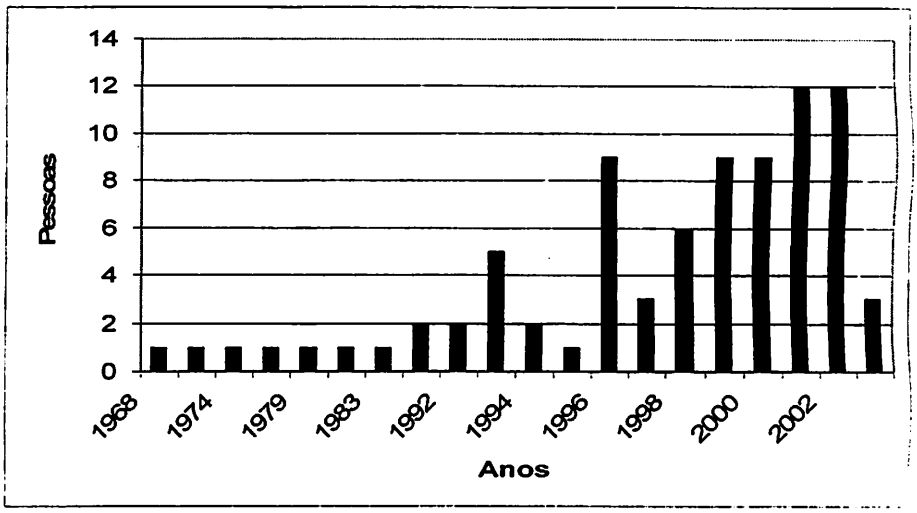


FIGURA 4.17 Associação temporal e vínculos profissionais com a UFLA
Fonte: Elaborado pela autora.

A Figura 4.18 apresenta o resultado de uma consulta para avaliar se há uma relação temporal entre o tempo de serviço das pessoas ligadas à UFLA e o ano de início de suas pesquisas cadastradas. Pelo gráfico percebe-se um padrão de comportamento, pois a grande maioria das pessoas cadastrou suas pesquisas mais recentes nos seus currículos. A função elaborada envolveu as tabelas de dados gerais das pessoas, as de atuações e a de atividades de pesquisa, resultando um total de 79 pesquisas.

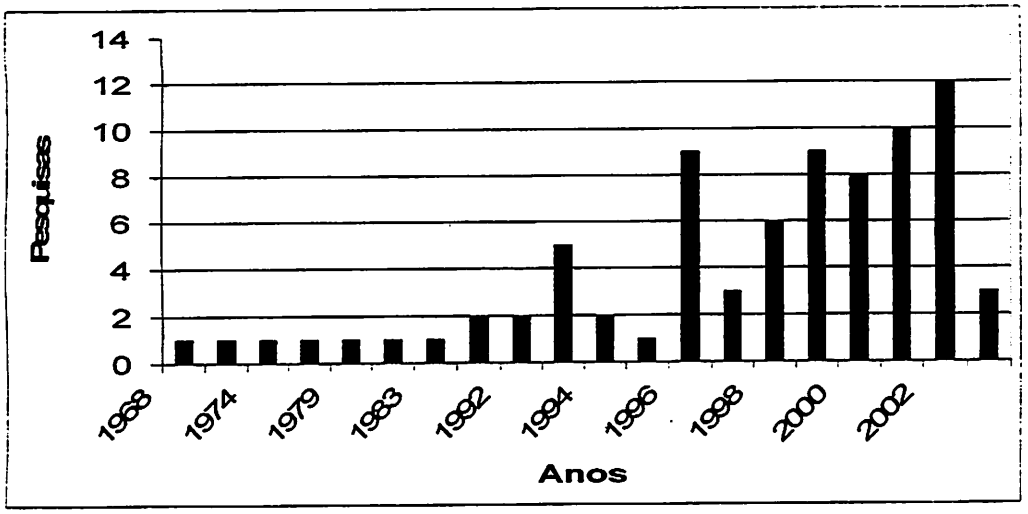


FIGURA 4.18 Associação temporal de pessoas e o ano de início de suas pesquisas
Fonte: Elaborado pela autora.

4.3.2.5 Análises de *clusters*

O exemplo a seguir faz a análise de um agrupamento (*cluster*) que inicialmente era desconhecido e surgiu a partir da consulta para verificar a duração, em anos, das pesquisas realizadas por pessoas da UFLA. O gráfico da Figura 4.19 mostra que, além das pesquisas que estão em andamento e não se pode afirmar a sua duração exata, a maioria das pesquisas dura entre 2 e 3 anos.

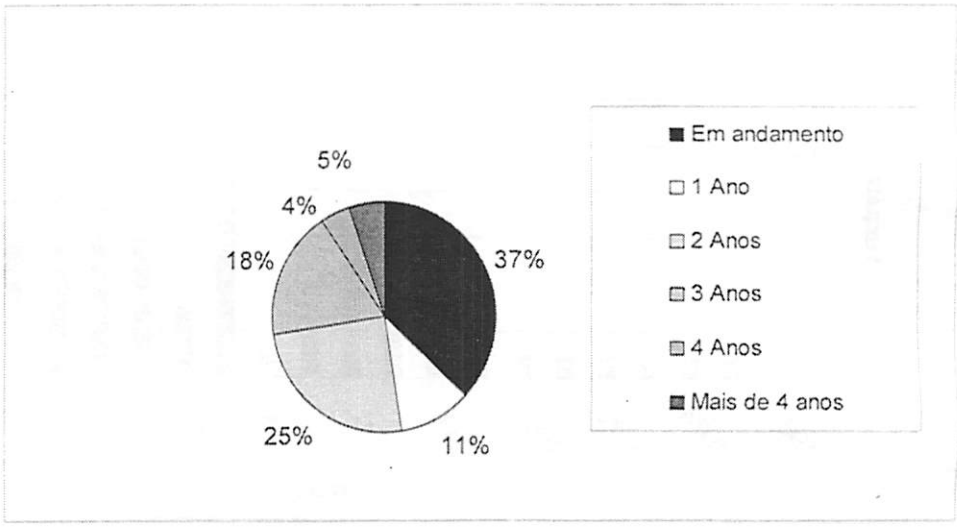


FIGURA 4.19 Agrupamento das pesquisas por tempo de duração
Fonte: Elaborado pela autora.

4.3.2.6 Análise de classificação e predição

A análise de classificação difere do agrupamento porque parte de grupos pré-definidos dos dados. Como as características dos dados extraídos da Plataforma Lattes não têm padrão definido, a tarefa de analisar os grupos já existentes tornou-se muito complexa, uma vez que faltava conhecimento da pesquisadora em agrupar, por exemplo, linhas ou áreas de pesquisa. Por este motivo apenas um exemplo será apresentado.

A consulta dividiu as atividades realizadas pelas pessoas da UFLA em três grupos: pesquisa, ensino e direção. O objetivo foi observar, dentre todos os currículos cadastrados, como foi a distribuição das publicações realizadas por pessoas enquanto estavam exercendo cada uma destas atividades. De um total de 101 publicações, o resultado obtido está ilustrado na Figura 4.20. Esta função envolveu três tabelas relacionadas às atividades e quatro tabelas relacionadas aos diversos tipos de publicações.

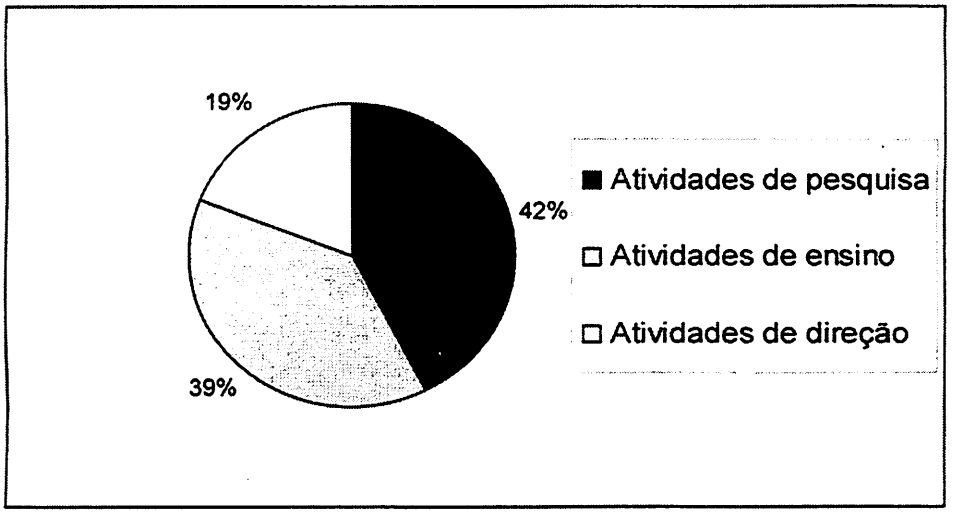


FIGURA 4.20 Relação entre tipos de atividades e publicações
Fonte: Elaborado pela autora.

Este resultado mostra que a maioria das publicações foi realizada enquanto as pessoas exerciam atividades de pesquisa; outra parte do total foi quando as pessoas exerciam atividades de ensino e, em menor número enquanto exerciam atividades de direção. Porém, estes agrupamentos não são disjuntos, ou seja, uma pessoa poderia estar ao mesmo tempo realizando diferentes tipos de atividades no momento da publicação. Mesmo assim, este é um resultado significativo, pois mostra claramente que, a depender do tipo de atividade em que a pessoa está envolvida, a quantidade de pub ~~ra~~ções que ela irá realizar será influenciada.

4.3.3 Análise dos resultados e sua aplicabilidade na UFLA

Todos os aspectos apresentados até o momento, tanto no que diz respeito à construção de indicadores para CT&I nas IES quanto no que diz respeito às políticas de gestão de CT&I na UFLA, estão relacionados diretamente com os dados semi-estruturados disponíveis na Plataforma Lattes. Em ambos os casos, uma grande importância é dada às informações contidas nesta Plataforma,

relativas à produção bibliográfica, à pós-graduação, às atividades realizadas pelas pessoas envolvidas em pesquisa nas IES, dentre outras. Porém, a grande quantidade de informação envolvida nestas áreas necessita claramente ser tratada de alguma forma para que o conhecimento seja criado e disseminado neste ambiente.

Mediante os trabalhos discutidos na Seção 2.4.1, os quais demonstram a importância dos indicadores de CT&I nas IES, um esforço deve ser realizado para criar tais indicadores para a UFLA. Os resultados até o momento apresentados demonstram que a Plataforma Lattes, uma vez devidamente atualizada, constitui-se de uma enorme fonte de informação para tal fim. Porém, para que os resultados reflitam a realidade da UFLA, de forma atualizada e segura, é necessário um grande esforço de todas as pessoas envolvidas com desenvolvimento científico, em preencher e ou atualizar seu Currículo Lattes. Antes dessa fase, entretanto, é importante que seja adotado um padrão para preenchimento do currículo, para que as funções deste grande conjunto de informações resultem num conhecimento consistente.

É importante mencionar o fato de que, até o ano de 2002, o currículo Lattes não era uma exigência institucional de peso, como é nos dias atuais. Foi só a partir daquele ano que ele tornou-se praticamente uma obrigatoriedade é a base para garantir financiamento de projetos, dentre outros.

Com relação aos resultados apresentados na Seção 4.3.2, os mesmos podem ser utilizados pelos mais diversos órgãos envolvidos diretamente com a produção científica da UFLA, tais como as coordenações dos diversos programas de Pós-Graduação da UFLA, a Reitoria, a Pró-Reitoria de Pesquisa e a Pró-Reitoria de Pós-Graduação, dentre outros. É importante que fique claro que tais funções foram elaboradas pela pesquisadora, com base apenas nas limitações apresentadas pelos próprios dados extraídos e na própria experiência de vida. Muitas outras consultas e funções podem ser construídas e aplicadas a

esta fonte de dados que agora, com este trabalho, encontra-se organizada de forma mais estruturada, facilitando a tarefa de analisá-la. Seria importante apresentar estes resultados às pessoas responsáveis pelos órgãos superiores da universidade, para que os mesmos possam contribuir para a construção de novas funções de *Data Mining*, com o objetivo de melhorar a ferramenta desenvolvida.

Alguns exemplos práticos da aplicabilidade destes resultados na UFLA poderiam ser:

- a partir da verificação da distribuição das atividades de ensino, de pesquisa e de direção, decisões poderiam ser tomadas para tentar não sobrecarregar as pessoas alocadas em determinados órgãos ou unidades, em detrimento de outros;
- analisar os diversos casos de pessoas que fogem ao padrão (*outliers*) dos demais, tentando verificar se este é ou não um bom comportamento, e se este deveria ser seguido, formando um novo padrão ou, ao contrário, ser evitado;
- a partir dos agrupamentos de pessoas que inicialmente não estão diretamente ligadas a nenhum departamento ou grupo de pesquisa, criar novas linhas ou áreas de pesquisas, que poderiam ser potencialmente melhor aproveitadas;
- a partir dos diversos padrões de comportamento observados nas informações que foram apresentadas, decisões podem ser tomadas não somente a curto prazo, mas também a longo prazo, pois é possível prever de forma segura prováveis comportamentos futuros;
- as diversas regras de associação que foram apresentadas mostram que dados que aparentemente não estão relacionados, na realidade, possuem aspectos em comum, que podem ser explorados;
- uma vez que a quantidade de publicações é uma medida de grande importância para as instituições de fomento de P&D nacionais, investir

nos órgãos da UFLA, que tradicionalmente produzem mais poderia ser uma boa estratégia, ou então, o inverso, investir naqueles órgãos que produzem pouco para tentar alavancar esta característica nos mesmos; etc.

O fato dos dados extraídos da Plataforma Lattes possuírem características de dados semi-estruturados torna a tarefa de interpretá-los ainda mais complexa. Outro aspecto importante diz respeito à possibilidade de aprofundar nos detalhes dos valores nominais que compõem o banco de dados criado. Isso depende da participação de pessoas que entendem das mais diversas áreas relativas às pesquisas realizadas na UFLA. Neste caso, uma sugestão seria a criação de comissões de especialistas nas diversas áreas, que pudessem avaliar tais informações. A partir dessa avaliação, os padrões poderiam ser criados.

5 CONCLUSÃO

O objetivo proposto neste trabalho foi o de construir e analisar uma ferramenta de *Data Mining*, como parte do processo de descoberta de conhecimento em banco de dados, para extrair conhecimento referente à produção científica das pessoas envolvidas com a UFLA, por meio dos dados extraídos da Plataforma Lattes. Para tanto, foi implementado um programa para transformar os dados semi-estruturados selecionados desta plataforma num banco de dados estruturado criado no Oracle. A partir daí, foi desenvolvida uma ferramenta automática de descoberta de conhecimento, utilizando a técnica de *Data Mining*, cujos resultados gerados foram analisados. Entende-se, portanto, que os objetivos foram alcançados.

Os resultados considerados mais expressivos e sua análise podem ser assim sintetizados:

- Com relação às limitações e aos problemas envolvendo os dados extraídos da Plataforma Lattes:
    - um dos grandes problemas encontrados para realizar a análise dos dados é a falta de padronização dos valores cadastrados;
    - outro problema refere-se ao próprio formato do currículo Lattes, que não deixa claro qual é a função de cada pessoa ligada à instituição;
    - poucas pessoas atualizam seus currículos Lattes periodicamente e, quando atualizam, a maioria dos currículos é preenchido de forma parcial.
- Dos primeiros resultados apresentados observando-se os números de atividades cadastradas, é interessante perceber que, ao preencherem seus currículos na Plataforma Lattes, dá-se maior prioridade às atividades de ensino e pesquisa do que às demais.

- Com relação às publicações:
    - percebe-se que a grande maioria delas pertence à grande área de Ciências Agrárias;
    - pessoas que não estão atuando na UFLA publicam mais do que quando estão; o fato de não estar atuando pode significar que possa estar fazendo pós-graduação e, por isso, tende a uma maior quantidade de produção e, conseqüentemente, de publicação. Por outro lado, ao estarem atuando na UFLA em atividades de ensino e direção, as pessoas têm menor disponibilidade de tempo para a produção de trabalhos em pesquisa, conseqüentemente, um número menor de pesquisas e publicações;
    - a média de publicações no exterior por pessoa é maior para aquelas que cursaram pós-graduação fora do Brasil;
    - a maioria das publicações foi realizada enquanto as pessoas exerciam atividades de pesquisa, seguidas pelas pessoas que exerciam atividades de ensino e, por fim, enquanto exerciam atividades de direção.
- É clara a importância dos indicadores de CT&I nas IES. Um esforço deve ser realizado para criar tais indicadores para a UFLA.
- A Plataforma Lattes, uma vez devidamente atualizada, constitui-se de uma enorme fonte de informação para a geração de conhecimento útil para a gestão das IES.

Diante dos resultados apresentados, pode-se perceber que, com esta ferramenta, é possível obter-se uma visão mais abrangente dos dados institucionais, pelo fato de ter sido disponibilizada uma grande quantidade de informações sobre a pesquisa científica da UFLA. Portanto, é possível iniciar 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.

Apesar de ter sido aplicada em uma área específica, a pesquisa científica na UFLA, o trabalho demonstrou como é possível também utilizar tecnologias da informação para auxiliar na gestão de conhecimento disponível nas Instituições de Ensino Superior. Diversos padrões e associações foram identificados por meio da aplicação da descoberta de conhecimento em banco de dados; porém, há muitas outras descobertas que ainda podem ser feitas aproveitando-se o banco de dados que foi criado.

Mesmo com algumas limitações, como a desatualização dos currículos e uma certa falta de padronização nos cadastros, este trabalho é uma iniciativa única no sentido de dar uma visão integrada das produções científica, tecnológica e bibliográfica de professores e pessoas ligadas à Universidade Federal de Lavras. Os dados e informações obtidos criaram um conjunto de informações, que pode servir de base para o processo de gestão da pesquisa científica nesta instituição.

Uma solução computacional para o problema da falta de padronização encontrada nos dados extraídos da Plataforma Lattes poderia ser a tentativa de realizar um estudo da semântica destes dados, por meio da construção de uma ontologia. Ontologia é a especificação explícita de uma conceitualização. Ela pode ser reconhecida como uma representação formal do vocabulário de termos de um domínio e define os termos necessários para descrever e representar uma área de conhecimento. A partir da representação e estrutura de um domínio é possível obter o conteúdo das informações.

Não se pode negar que as universidades devem fazer uso da gestão do conhecimento para auxiliar a tomada de decisão na busca de uma qualidade cada vez melhor na prestação de seus serviços. No caso da UFLA, ações organizacionais podem ser tomadas, destinadas à gestão do conhecimento

interno e externo à instituição. Para isso, sugere-se a criação de uma comissão de especialistas das mais diversas áreas de conhecimento da UFLA, com o objetivo de elaborar um documento que defina um padrão para o preenchimento e a atualização dos currículos Lattes pelos envolvidos com a pesquisa científica na UFLA. A partir deste trabalho, um outro esforço pode ser realizado para a elaboração de indicadores de CT&I destinados à UFLA, tendo como base aqueles discutidos na Seção 2.4.1.

Por fim, pode-se dizer que 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, como: (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, gerando descoberta de novas informações e novo conhecimento, trazendo melhorias para a ferramenta desenvolvida; (3) criação de uma comissão que elabore normas para o preenchimento e atualização dos currículos Lattes das pessoas envolvidas com a pesquisa científica na UFLA; (4) criação de indicadores de CT&I para a UFLA, com o objetivo de auxiliar a elaboração de novas políticas de gestão; (5) a aplicação da ferramenta desenvolvida nos currículos atualizados, assim que estes estejam disponíveis na Plataforma lattes, e comparação dos novos resultados obtidos com os resultados obtidos neste trabalho e (6) aplicação desta ferramenta em outras instituições de ensino superior, com o objetivo de comparar seus resultados com aqueles obtidos na UFLA.

Poucos trabalhos vêm sendo realizados sobre a Plataforma Lattes com o objetivo de tentar melhor utilizar suas informações. Acredita-se que este trabalho seja uma contribuição que pode servir para melhorar a própria plataforma, tão amplamente utilizada pelos pesquisadores nacionais.

Este trabalho foi apenas uma etapa de um processo maior de desenvolvimento do conhecimento, que pretende servir de apoio à tomada de decisão, devido à possibilidade de que, no futuro, sejam criados indicadores para serem aplicados às instituições de ensino superior. Os resultados comparativos do uso destes indicadores trariam grande contribuição à gestão da política científica e tecnológica e ao aperfeiçoamento do sistema de ensino superior brasileiro.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ABITEBOUL, S. Querying semistructured data. In: INTERNATIONAL CONFERENCE ON DATABASE THEORY, ICDT'97, 1997. Greece, 1997. **Proceedings of international conference on database theory.** Greece: Delphi, 1997. v. 1186, p. 1-18.

ADRIAANS, P.; ZANTINGE, D. **Data mining**. Harlow: Addison-Wesley, 1996. 158p.

ALVARENGA, R. et al. Gestão de conhecimento para ensino e pesquisa: o modelo da UCB. In: CONGRESSO ANUAL DA SOCIEDADE BRASILEIRA DE GESTÃO DO CONHECIMENTO, São Paulo, 2002. **Anais do congresso anual da sociedade brasileira de gestão do conhecimento.** Disponível em: <http://www.cori.rei. unicamp.br>. Acesso em: 10 out. 2004.

AMO, S. **Curso de Data Mining**: programa de mestrado em ciência da computação. Uberlândia: Universidade Federal de Uberlândia, 2003. Disponível em: <http://www.deamo.prof.ufu.br/CursoDM.html.>. Acesso em: 05 jul. 2004.

AMO, S., Mineração de dados. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 24, 2004, Salvador. **Anais do 23 congresso da sociedade brasileira de computação .** Salvador: Sociedade Brasileira de Computação, 2004. v.2. p. 195-238.

BONOMA, T.V. Case research in marketing: opportunities, problems, and process. **Journal of Marketing Research**, v.22, 209p., maio, 1985.

BRASIL. Ministério da Ciência e Tecnologia. **Indicadores de ciência e tecnologia**. 2001. Disponível em: <http://www.mct.gov.br>. Acesso em: 03 jul. 2004.

BRESSAN, F. O método do estudo de caso. **Administração On Line**, São Paulo, v.1, n.1, 2000. (FEA-USP) Disponível em: <http://www.fecap.br/adm_online>. Acesso em: 10 out. 2004.

CARVALHO, R.B. **Aplicações de softwares de gestão do conhecimento:** tipologia e usos. 2000. Dissertação (Mestrado em Ciência da Computação)- Universidade Federal de Minas Gerais, Belo Horizonte.

COELHO, M.I.M. Gestão de C&T: o que é. In: _____. **Gestão de C&T**: planejamento de pesquisa e captação de recursos. 2002. Disponível em: <http://netpage.em.com.br/mines/pesquisa.htm>. Acesso em: 04 out. 2004.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. **Conselho nacional de desenvolvimento científico e tecnológico**. Disponível em: <http://www.cnpq.br/>. Acesso em: 10 abr. 2004.

DECKER, K.; FOCARDI, S. **Technological overview**: a report on data mining. CSCS – Swiss National Supercomputing Center, Techinical Report, Zurique, 1995. Disponível em: <ftp://ftp.cscs.ch/pub/CSCS/> Acesso em: 17 mar. 2004.

DEITEL, H.M. et al. **XML como programar**. Porto Alegre: Bookman, 2003.

ELMASRI, R.; NAVATHE, S.B. **Sistemas de banco de dados:** fundamentos e aplicações. 3.ed. Rio de Janeiro: LTC, 2002.

FAYYAD, U.M. et al. From data mining to knowledge discovery: an overview. In: **Advances in knowledge discovery and data mining**. California: AAAI/The MIT, 1996. p.1-34.

GOODE, W.J.; HATT, P.K. **Métodos em pesquisa social**. 3 ed. São Paulo: Cia Editora Nacional, 1969.

GRUPO STELA. **Documento LMPLCurriculo**. Florianópolis: Universidade Federal de Santa Catarina, 2002. Disponível em: <http://lattes.cnpq.br/lmpl/Gramaticas/Curriculo/XSD/Documentacao>. Acesso em: 10 fev. 2004.

GRUPO STELA, **Lattes Extrator**. Universidade Federal de Santa Catarina, Florianóplois, 2002. Disponível em: <http://lattes.cnpq.br /lattesextrator/>. Acesso em: 07/10/2004.

GROSSMAN, R.L, HORNICK, M., MEYER G., *Emerging KDD Standards*. In: ***Communications of the ACM***, Special Issue on Data Mining, 2002.

HAYASHI, M.C.P.I. Os indicadores de C&T como ferramenta de gestão da informação científica e tecnológica no contexto universitário. In: CONGRESSO ANUAL DA SOCIEDADE BRASILEIRA DE GESTÃO DO CONHECIMENTO, São Paulo, 2002. **Anais do congresso anual da sociedade brasileira de gestão do conhecimento.** São Paulo: SBGC, 2002. 16p.

HAYASHI, M.C.P.I. et al. **Ciência, tecnologia e inovação no pólo tecnológico de São Carlos.** São Carlos: Universidade Federal de São Carlos/Departamento de Ciência da Informação, 2004. Disponível em: <http://www.cori.rei.unicamp.br/IAU>. Acesso em: 02 out. 2004.

KING, D. **Numerical machine learning.** Georgia: Tech College of Computing, 2003. Disponível em: <http://www.cc.gatech.edu/kingd/datamine/datamine.html>. Acesso em: 22 mar. 2004.

KOCHE, J.C. **Fundamentos de metodologia científica**: teoria da ciência e prática de pesquisa. 14.ed. Petrópolis: Vozes, 1997.

KROGH, G.V.; ICHIJO, K.; NONAKA, I. **Facilitando a criação de conhecimento**. Rio de Janeiro, Campus, 2001.

LAUDON, K.C.; JANE, P. **Gerenciamento de sistema de informação**, 3.ed. Rio de Janeiro: LTC, 1999.

MACIAS-CHAPULA, C.A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ciência da Informação**, Brasília, v.27, n.2, p.134-140, maio/ago. 1998.

MORZY T.; WOJCIECHOWSKI M.; ZAKRZEWICZ M. Data mining support in database management systems. In: DAWAK CONFERENCE, POZNAN UNIVERSITY OF TECHNOLOGY, 2nd, 2000. **Proceedings...** Disponível em: <http://www.cs.put. poznan.pl/mwojciechowski/papers>. Acesso em: 21abr. 2004

MOXTON, B. Defining data mining. **DBMS Data Warehouse Supplement Site**, 2004. Disponível em: <http://www.dbms.mfi.com/ 9608d53.html> Consultado em 20 mar. 2004.

NAVEGA S. Princípios essenciais do data mining. In: INFOIMAGEM, 2002, Cenadem, novembro, 2002. **Anais do infoimagem.** Disponível em: <http://www.intelliwise.com/snavega>. Acesso em: 14 mar. 2004.

ORACLE. **Data mining:** an Oracle white paper, 2004. Disponível em: <http://www.oracle.com>. Acesso em 29 abr. 2004.

PARRINI, E. **Gestão do conhecimento no suporte à decisão OLAP**. 2002. Dissertação (Mestrado em Informática)-Universidade Federal do Rio de Janeiro/ COOPE, Rio de Janeiro.

PIROLLA, V.S. **A proposição de uma ferramenta de apoio ao mapeamento do conhecimento em uma organização**. 2002. Dissertação (Mestrado em Informática)-Universidade Federal do Rio de Janeiro/COPPE, Rio de Janeiro.

UNIVERSIDADE FEDERAL DE LAVRAS. Pró-Reitoria de Pesquisa da Ufla. Apresenta informações sobre a pós-graduação da UFLA. Disponível em: <http://www.prp.ufla.br>. Acesso em: 20 mar. 2004.

QUONIAM, L. et al. Inteligência obtida pela aplicação de data mining em base de teses francesas sobre o Brasil. **Ciência da Informação**, Brasília, v.30, n.2, p.20-28, maio/ago. 2001.

RAMALHO, J.A. **XML A informação na medida certa**. São Paulo: Berkeley, 2002. (Série Ramalho: Teoria e Prática).

SANTOS, M.F. **Descoberta de conhecimento em bases de dados**. Portugal: Universidade do Minho/Departamento de Sistemas de Informação, 2002. Disponível em: <http://piano.dsi.uminho.pt> Acesso em: 20 mar. 2004.

SCHWARTZMAN, S. **Organização e desempenho da pesquisa científica no Brasil (Projeto ICSOPRU)**, Rio de Janeiro, 1985. Disponível em: <http://www.schwartzman.org.br/simon>. Acesso em: 10 fev. 2004.

TACHIZAWA, T.; ANDRADE, R.O.B. **Gestão de instituições de ensino**. 3.ed. Rio de Janeiro: FGV, 2002.

TARAPANOFF, K. (Org.). **Inteligência organizacional e competitiva**. Brasília: Universidade de Brasília, 2001.

W3C, **Transformações XSL (XSLT)**. World Wide Web Consortium. Recomendação de 16 de novembro de 1999. Versão 1.0. Disponível em: <http://www.amtechs.com/w3c>. Acesso em: 04 maio 2004.

YIN, R.K. **Case study research:** design and methods. USA: Sage, 1989.

7 ANEXOS

7.1 ANEXO A – Tabelas do banco de dados extraído da Plataforma Lattes

1. 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}

2. END_PROF = {cod_pessoa, nome_inst, nome_unid, nome_orgao, pais, uf, complemento, bairro, cidade, cxpostal, cep, ddd, telefone, ramal, fax, email, home_page}

3. GRADUACAO = {cod_pessoa, seq_form, nivel, titulo_conclusao, nome_orientador, nome_inst, nome_curso, status_curso, ano_inicio, ano_conclusao, flag_bolsa, nome_agencia}

4. TIPO_POSGRAD = {cod_tipo, tipo}

5. POS_GRADUACAO = {cod_posgrad, cod_tipo, cod_pessoa, seq_form, nivel, nome_inst, nome_curso, nome_area, status_curso, ano_inicio, ano_conclusao, flag_bolsa, nome_agencia, ano_obtencao_titulo, titulo_dissertacao, nome_orientador}

6. PALAVRAS_CHAVE_POSGRAD = {cod_posgrad, palavra1, palavra2, palavra3, palavra4, palavra5, palavra6}

7. AREA_POSGRAD = {cod_posgrad, cod_area, grande_area, area, sub_area, especialidade}

8. SETOR_ATIVIDADE_POSGRAD = {cod_posgrad, setor1, setor2, setor3}

9. ATUACOES = {cod_atuacao, cod_pessoa, tipo}

10. 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}

11. ATIVIDADE_DIRECAO = {cod_atuacao, seq_funcao, flag_periodo, mes_inicio, ano_inicio, mes_fim, ano_fim, nome_orgao, nome_unid, formato_cargo, cargo_funcao}

12. ATIVIDADE_PESQUISA = {cod_pd, cod_atuacao, seq_funcao, flag_periodo, mes_inicio, ano_inicio, mes_fim, ano_fim, nome_orgao, nome_unid}

13. LINHA_PESQUISA = {cod_pd, seq_linha_pesq, titulo_linha, flag_linha_ativa, objetivos_linha}

14. ATIVIDADE_ENSINO = {cod_ensino, cod_atuacao, seq_funcao, flag_periodo, tipo_ensino, mes_inicio, ano_inicio, mes_fim, ano_fim, nome_orgao, nome_curso}

15. DISCIPLINA = {cod_ensino, seq_especificacao, comentário}

16. ATUACAO_PROF = {cod_atuacao, cod_pessoa, nome_inst, seq_ativ}

17. ATIVIDADES_EXTENSAO = {cod_atuacao, seq_funcao, mes_inicio, ano_inicio, mes_fim, ano_fim, nome_orgao, nome_unid, atividade_extensao}

18. SERVICOS_TECNICOS = {cod_atuacao, seq_funcao, flag_periodo, mes_inicio, ano_inicio, mes_fim, ano_fim, nome_orgao, nome_unid, servico_realizado}

19. TREINAMENTOS_MINISTRADOS = {cod_treinamento, cod_atuacao, seq_funcao, flag_periodo, mes_inicio, ano_inicio, mes_fim, ano_fim, nome_orgao, nome_unid}

20. TREINAMENTO = {cod_treinamento, seq_especificacao, observação}

21. OUTRAS_ATIVIDADES = {cod_atuacao, seq_funcao, flag_periodo, mes_inicio, ano_inicio, mes_fim, ano_fim, nome_orgao, nome_unid, atividade_realizada}

22. TRABALHOS_EM_EVENTOS = {cod_trab, cod_pessoa, seq_producao, natureza, titulo, ano, pais_evento, idioma, meio_divulgacao, home_page, flag_relevancia}

23. DETALHAMENTO_TRAB = {cod_trab, classificacao_evento, nome_evento, cidade_evento, ano_realizacao, titulo_anais, volume, fascículo, serie, pagina_inicial, pagina_final, isbn, editora, cidade_editora}

24. AUTORES_TRAB = {cod_trab,nome_autor,nome_citacao, ordem_autoria}

25. PALAVRAS_CHAVE_TRAB = {cod_trab, palavra1, palavra2, palavra3, palavra4, palavra5, palavra6}

26. AREA_DO_CONHECIMENTO_TRAB = {cod_trab, cod_area, grande_area, area, sub_area, especialidade}

27. INFORMACOES_ADICIONAIS_TRAB = {cod_trab, informações}

28. ARTIGOS_PUBLICADOS = {cod_artigo, cod_pessoa, seq_producao, natureza, titulo, ano, pais_publicacao, idioma, meio_divulgacao, home_page, flag_relevancia}

29. DETALHAMENTO_ARTIGO = {cod_artigo, titulo_periodico _revista, issn, volume, fascículo, serie, pagina_inicial, pagina_final, local_publicacao}

30. AUTORES_ARTIGO = {cod_artigo, nome, nome_citacao, ordem_autoria}

31. PALAVRAS_CHAVE_ARTIGO = {cod_artigo, palavra1, palavra2, palavra3, palavra4, palavra5, palavra6}

32. AREA_CONHECIMENTO_ARTIGO = {cod_artigo, cod_area, grande_area, área, sub_area, especialidade}

33. SETOR_ATIVIDADE_ARTIGO = {cod_artigo, setor1, setor2, setor3}

34. INFORMACOES_ADICIONAIS_ARTIGO = {cod_artigo, informações}

35. OUTRA_PROD_BIBLI = {cod_producao, cod_pessoa, seq_producao, natureza, titulo, ano, pais_publicacao, idioma, meio_divulgacao, home_page, flag_relevancia}

36. DETALHAMENTO_OUTRA_PROD = {cod_producao, editora, cidade_editora, numero_paginas, issn_isbn}

37. AUTORES_OUTRA_PROD = {cod_producao, nome_autor, nome_citacao, ordem_autoria}

38. TRABALHOS_TECNICOS = {cod_trab_tec, cod_pessoa, seq_producao, natureza, titulo, ano, pais, idioma, meio_divulgacao, home_page, flag_relevancia}

39. DETALHAMENTO_TRAB_TEC = {cod_trab_tec, finalidade, duracao_meses, numero_paginas, disponibilidade, inst_financiadora, cidade}

40. AUTORES_TRAB_TEC = {cod_trab_tec, nome_autor, nome_citacao, ordem_autoria}

41. PALAVRAS_CHAVE_TRAB_TEC = {cod_trab_tec, palavra1, palavra2, palavra3, palavra4, palavra5, palavra6}

42. AREA_CONHECIMENTO_TRAB_TEC = {cod_trab_tec, cod_area, grande_area, area, sub_area, especialidade}

43. SETOR_ATIVIDADE_TRAB_TEC = {cod_trab_tec, setor1, setor2, setor3}

44. INF_ADICIONAIS_TRAB_TEC = {cod_trab_tec, informações}

45. ORIENTACOES_CONCLUIDAS = {cod_orientacao, cod_pessoa, seq_producao, natureza, tipo, titulo, ano, pais, idioma, home_page, flag_relevancia}

46. DETALHAMENTO_ORIENTACOES = {cod_orientacao,nome_orientado, nome_inst, nome_curso, flag_bolsa, nome_agencia, tipo_orientacao, numero_paginas}

47. PALAVRAS_CHAVE_ORIENTACOES = {cod_orientacao, palavra1, palavra2, palavra3, palavra4, palavra5, palavra6}

48. INF_ADICIONAIS_ORIENTACOES = {cod_orientacao, informações}

49. PARTICIPACAO_BANCA = {cod_participacao, cod_pessoa, seq_producao, natureza, tipo, titulo, ano, pais, idioma, home_page}

50. DETALHAMENTO_BANCA = {cod_participacao, nome_candidato, nome_inst, nome_curso}

51. PARTICIPANTE_BANCA = {cod_participacao, nome_participante, nome_citacao, ordem_participante}

52. INF_ADICIONAIS_BANCA = {cod_producao, informacoes}

53. INF_ADICIONAIS_OUTRA_PROD = {cod_producao, informacoes}

54. AREA_CONHECIMENTO_OUTRA_PROD = {cod_producao, cod_area, grande_area, area, sub_area, especialidade}

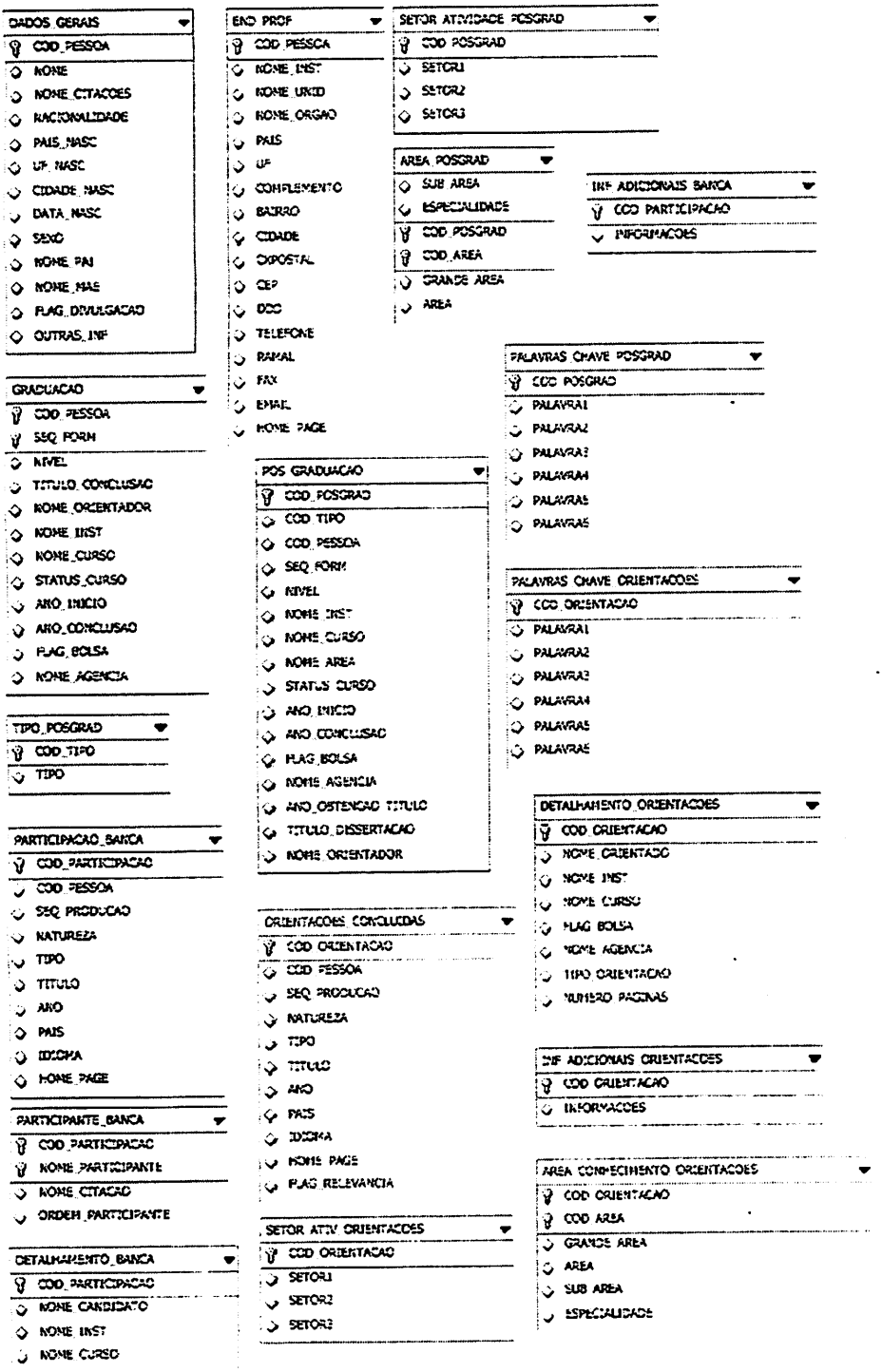
55. PALAVRAS_CHAVE_OUTRA_PROD = {cod_producao, palavra1, palavra2, palavra3, palavra4, palavra5, palavra6}

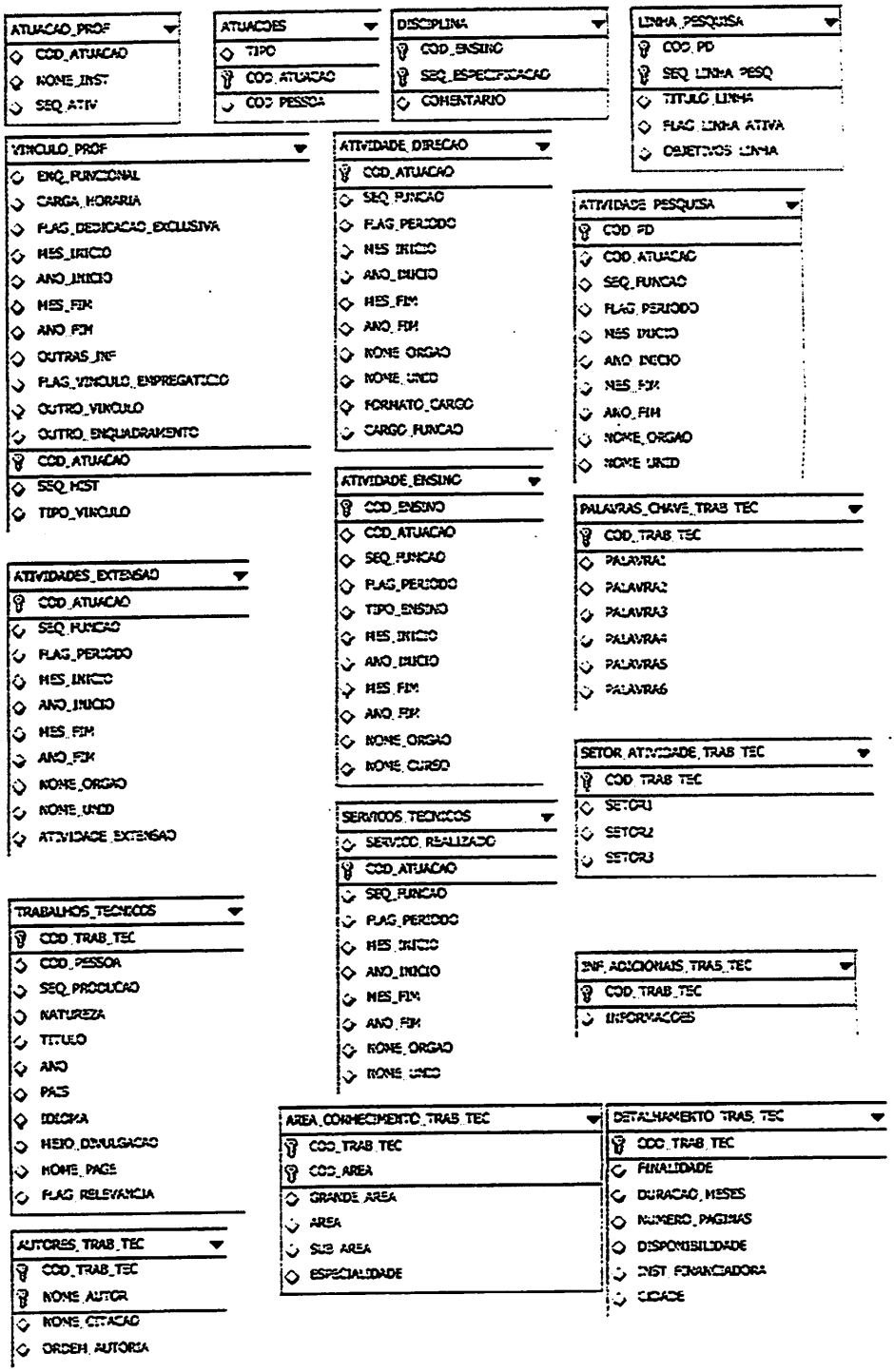
56. SETOR_ATIV_OUTRA_PROD = {cod_producao, setor1, setor2, setor3}

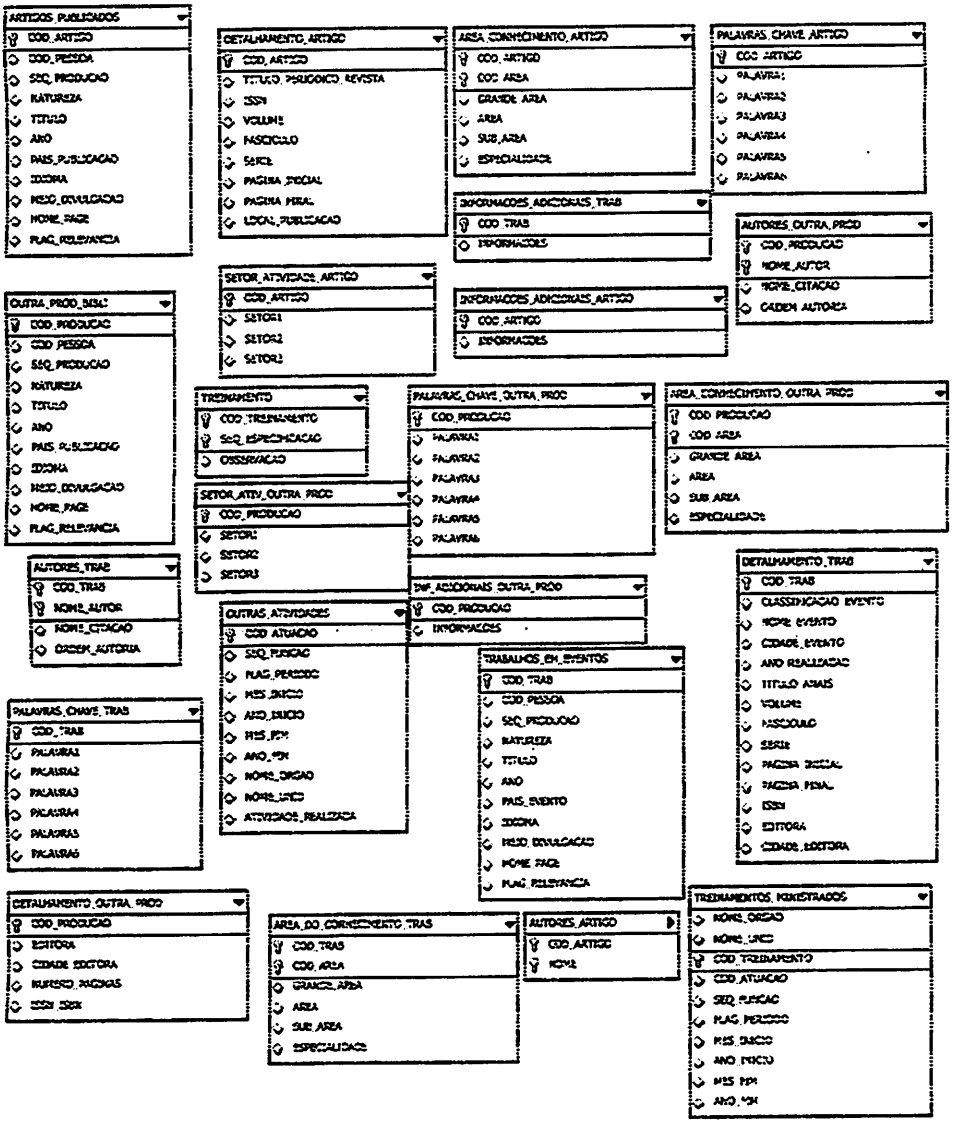
57. SETOR_ATIV_ORIENTACOES = {cod_orientacao, setor1, setor2, setor3}

58. AREA_CONHECIMENTO_ORIENTACOES = {cod_orientacao,cod_area, grande_area, area, sub_area, especialidade}

7.2 ANEXO B – Modelo Entidade Relacionamento (ER) do banco de dados dos currículos Lattes




ATUACAO_PROF
COD_ATUACAO
NOME_INST
SEQ_ATIV
ATUACOES
TIPO
COD_ATUACAO
COD_PESSOA
DISCIPLINA
COD_ENSINO
SEQ_ESPECIFICACAO
COMENTARIO
LINHA_PESQUISA
COD_PD
SEQ_LINHA_PESQ
TITULO_LINHA
FLAG_LINHA_ATIVA
OBJETIVOS_LINHA
VINCULO_PROF
ENQ_FUNCIONAL
CARGA_HORARIA
FLAG_DEDICACAO_EXCLUSIVA
MES_INICIO
ANO_INICIO
MES_FIM
ANO_FIM
OUTRAS_INF
FLAG_VINCULO_EMPREGATICIO
OUTRO_VINCULO
OUTRO_ENQUADRAMENTO
COD_ATUACAO
SEQ_INST
TIPO_VINCULO
ATIVIDADE_DIRECAO
COD_ATUACAO
SEQ_FUNCAO
FLAG_PERIODO
MES_INICIO
ANO_INICIO
MES_FIM
ANO_FIM
NOME_ORGAO
NOME_UNID
FORMATO_CARGO
CARGO_FUNCAO
ATIVIDADE_PESQUISA
COD_PD
COD_ATUACAO
SEQ_FUNCAO
FLAG_PERIODO
MES_INICIO
ANO_INICIO
MES_FIM
ANO_FIM
NOME_ORGAO
NOME_UNID
ATIVIDADE_ENSINO
COD_ENSINO
COD_ATUACAO
SEQ_FUNCAO
FLAG_PERIODO
TIPO_ENSINO
MES_INICIO
ANO_INICIO
MES_FIM
ANO_FIM
NOME_ORGAO
NOME_CURSO
PALAVRAS_CHAVE_TRAB_TEC
COD_TRAB_TEC
PALAVRA1
PALAVRA2
PALAVRA3
PALAVRA4
PALAVRA5
PALAVRA6
ATIVIDADES_EXTENSAO
COD_ATUACAO
SEQ_FUNCAO
FLAG_PERIODO
MES_INICIO
ANO_INICIO
MES_FIM
ANO_FIM
NOME_ORGAO
NOME_UNID
ATIVIDADE_EXTENSAO
SETOR_ATIVIDADE_TRAB_TEC
COD_TRAB_TEC
SETOR1
SETOR2
SETOR3
SERVICOS_TECNICOS
SERVICO_REALIZADO
COD_ATUACAO
SEQ_FUNCAO
FLAG_PERIODO
MES_INICIO
ANO_INICIO
MES_FIM
ANO_FIM
NOME_ORGAO
NOME_UNID
TRABALHOS_TECNICOS
COD_TRAB_TEC
COD_PESSOA
SEQ_PRODUCAO
NATUREZA
TITULO
ANO
PAIS
IDIOMA
MEIO_DIVULGACAO
NOME_PAGE
FLAG_RELEVANCIA
INF_ADICIONAIS_TRAB_TEC
COD_TRAB_TEC
INFORMACOES
AUTORES_TRAB_TEC
COD_TRAB_TEC
NOME_AUTOR
NOME_CITACAO
ORDEM_AUTORIA
AREA_CONHECIMENTO_TRAB_TEC
COD_TRAB_TEC
COD_AREA
GRANDE_AREA
AREA
SUB_AREA
ESPECIALIDADE
DETALHAMENTO_TRAB_TEC
COD_TRAB_TEC
FINALIDADE
DURACAO_MESES
NUMERO_PAGINAS
DISPONIBILIDADE
INST_FINANCIADORA
CIDADE


ARTIGOS_PUBLICADOS
COD_ARTIGO
COD_PESSOA
SEQ_PRODUCAO
NATUREZA
TITULO
ANO
PAIS_PUBLICACAO
IDIOMA
MEIO_DIVULGACAO
HOME_PAGE
FLAG_RELEVANCIA
DETALHAMENTO_ARTIGO
COD_ARTIGO
TITULO_PERIODICO_REVISTA
ISSN
VOLUME
FASCICULO
SERIE
PAGINA_INICIAL
PAGINA_FINAL
LOCAL_PUBLICACAO
AREA_CONHECIMENTO_ARTIGO
COD_ARTIGO
COD_AREA
GRANDE_AREA
AREA
SUB_AREA
ESPECIALIDADE
PALAVRAS_CHAVE_ARTIGO
COD_ARTIGO
PALAVRA1
PALAVRA2
PALAVRA3
PALAVRA4
PALAVRA5
PALAVRA6
INFORMACOES_ADICIONAIS_TRAB
COD_TRAB
INFORMACOES
AUTORES_OUTRA_PROD
COD_PRODUCAO
NOME_AUTOR
NOME_CITACAO
ORDEM_AUTORIA
SETOR_ATIVIDADE_ARTIGO
COD_ARTIGO
SETOR1
SETOR2
SETOR3
INFORMACOES_ADICIONAIS_ARTIGO
COD_ARTIGO
INFORMACOES
OUTRA_PROD_BIBLIO
COD_PRODUCAO
COD_PESSOA
SEQ_PRODUCAO
NATUREZA
TITULO
ANO
PAIS_PUBLICACAO
IDIOMA
MEIO_DIVULGACAO
HOME_PAGE
FLAG_RELEVANCIA
TREINAMENTO
COD_TREINAMENTO
SEQ_ESPECIFICACAO
OBSERVACAO
PALAVRAS_CHAVE_OUTRA_PROD
COD_PRODUCAO
PALAVRA1
PALAVRA2
PALAVRA3
PALAVRA4
PALAVRA5
PALAVRA6
AREA_CONHECIMENTO_OUTRA_PROD
COD_PRODUCAO
COD_AREA
GRANDE_AREA
AREA
SUB_AREA
ESPECIALIDADE
SETOR_ATIV_OUTRA_PROD
COD_PRODUCAO
SETOR1
SETOR2
SETOR3
AUTORES_TRAB
COD_TRAB
NOME_AUTOR
NOME_CITACAO
ORDEM_AUTORIA
INF_ADICIONAIS_OUTRA_PROD
COD_PRODUCAO
INFORMACOES
DETALHAMENTO_TRAB
COD_TRAB
CLASSIFICACAO_EVENTO
NOME_EVENTO
CIDADE_EVENTO
ANO_REALIZACAO
TITULO_ANAIS
VOLUME
FASCICULO
SERIE
PAGINA_INICIAL
PAGINA_FINAL
ISSN
EDITORA
CIDADE_EDITORA
OUTRAS_ATIVIDADES
COD_ATUACAO
SEQ_FUNCAO
FLAG_PERIODO
MES_INICIO
ANO_INICIO
MES_FIM
ANO_FIM
NOME_ORGAO
NOME_UNID
ATIVIDADE_REALIZADA
TRABALHOS_EM_EVENTOS
COD_TRAB
COD_PESSOA
SEQ_PRODUCAO
NATUREZA
TITULO
ANO
PAIS_EVENTO
IDIOMA
MEIO_DIVULGACAO
HOME_PAGE
FLAG_RELEVANCIA
PALAVRAS_CHAVE_TRAB
COD_TRAB
PALAVRA1
PALAVRA2
PALAVRA3
PALAVRA4
PALAVRA5
PALAVRA6
DETALHAMENTO_OUTRA_PROD
COD_PRODUCAO
EDITORA
CIDADE_EDITORA
NUMERO_PAGINAS
ISSN_ISBN
AREA_DO_CONHECIMENTO_TRAB
COD_TRAB
COD_AREA
GRANDE_AREA
AREA
SUB_AREA
ESPECIALIDADE
AUTORES_ARTIGO
COD_ARTIGO
NOME
TREINAMENTOS_REGISTRADOS
NOME_ORGAO
NOME_UNID
COD_TREINAMENTO
COD_ATUACAO
SEQ_FUNCAO
FLAG_PERIODO
MES_INICIO
ANO_INICIO
MES_FIM
ANO_FIM

7.3 ANEXO C – Exemplos de dados cadastrados no Currículo Lattes de forma redundante ou errônea

72 cargos distintos cadastrados nos currículos de pessoas ligadas à UFLA:

1. Administradora de Banco de Dados
2. Analista de Sistemas
3. Assessor Administrativo do DEG
4. CHEFE DE DEPARTAMENTO
5. COORDENADOR DE MS. EM CIENCIA DOS ALIMENTOS
6. Chefe Adjunto Técnico
7. Chefe de departamento interinamente
8. Chefe do Departamento
9. Chefe do Departamento de Ciência da Computação
10. Chefe-Geral
11. Conselheiro suplente
12. Coordenador
13. Coordenador Finaceiro do Programa de Apoio Tecnológico
14. Coordenador Geral de Pós-Graduação
15. Coordenador da incubadora UFLAtec
16. Coordenador da Área de Sistemas de Produção
17. Coordenador de Laboratório
18. Coordenador de Programa
19. Coordenador do Convênio ESAL/FEBEN em Itanhandú, MG
20. Coordenador do Curso de Pós-Graduação em Agronomia/Fisiologia Vegetal Port. 326 de 16/7/1996
21. Coordenador do Laboratório de Água e Solo
22. Coordenador do curso de Eng. Agrícola
23. Coordenador local do Programa de Desenvolvimento de Gado Leiteiro financiado pela CIDA, Canadá.
24. Coordenadora da àrea de Sistemas de Ensino de Graduação
25. Diretor Comercial
26. Diretor Financeiro
27. Diretor Presidente
28. Gerente da Fazenda Experimental de Caldas
29. Gerente o objetivo estratégico Parcerias Externas.
30. Implantação e Co-responsável do Laboratório de Cultura de Tecidos Vegetais
31. MEMBRO DA EGREGIA CONGREGACAO DA UFLA NO PERIODO DE 13/09/93 A 28/02/95
32. MEMBRO DO CONSELHO CURADOR

33. Membro Representante da Classe Adjunta junto à CPPD
34. Membro da Comissão de Assuntos Acadêmicos
35. Membro da Comissão de Ensino de Graduação
36. Membro da Comissão de Licitação
37. Membro de Colegiado de Curso de Pósgraduação em Agronomia/Fisiologia Vegetal-Port. 003 de 22/1/1992
38. Membro de Colegiado de Curso de Pósgraduação em Agronomia/Fisiologia Vegetal-Port. 006 de 19/02/1988
39. Membro de Colegiado do Curso de Engenharia Florestal -Of. 046/93/DCF
40. Membro de Colegiado do Curso de Pós-Graduação em Agronomia/Fitotecnia Port.PRPG 109 de 12/6/2000
41. Membro de Comissão Temporária
42. Membro de Conselho de Ensino Pesquisa e Extensão - CEPE
43. Membro de comissão permanente
44. Membro de comissão permanente da BC
45. Membro do Colegiado do Curso de Pós Graduação em Solos e Nutrição de Plantas
46. Membro do Colegiado do Curso de Pós-Graduação em Agronomia/Fisiologia Vegetal Port.PRPG 111 de 12/6/2000
47. Membro do Conselho Universitário
48. Membro do Conselho de Biblioteca
49. Membro do Conselho de Ensino, Pesquisa e Extensão
50. Prefeito do Campus Universitário
51. Presidente da CPPD
52. Presidente da Comissão Organizadora do V Congresso Brasileiro de Fisiologia Vegetal
53. Presidente da Comissão de Estágios
54. Presidente de Comissão Permanente de Pessoal Docente -CPPD
55. Programadora de Sistemas
56. Responsável pela Chefia do Departamento de Engenharia Rural
57. Responsável pela Estação Gráfica do DEG
58. Responsável pelo Lab. de Fotointerpretação
59. Secretário Executivo da Comissão Técnica do Programa Sistemas de Produção Animal coordenado opela Embrapa
60. Secretário Executivo do Programa Fruticultura
61. Secretário da Sociedade Brasileira de Fisiologia Vegetal
62. Sub-Chefe do Departamento de Biologia
63. Sub-coordenador de Curso de Pósgraduação em Agronomia/Fisiologia Vegetal
64. Sub-coordenador de programa

65. Sub-coordenadora do Curso de Mestrado Profissional em Gestão em Negócios
66. Subchefe do Departamento de Engenharia Rural
67. Supervisor de Informática do DEG
68. Supervisor do Convênio Interinstitucional UFLA/CEFET-PR
69. Supervisora da Divisão de Desenvolvimento
70. Supervisora do Setor de Programação
71. Vice-coordenador
72. Vice-presidente da COPEVE

46 órgãos diferentes cadastrados nos currículos de pessoas ligadas à UFLA:

1. Agricultura Alternativa
2. Botanic Departament
3. Centro Nacional de Pesquisa de Gado de Leite
4. Centro Nacional de Pesquisa de Soja
5. Centro Nacional de Pesquisa de Solos
6. Centro Tecnológico do Sul de Minas
7. Centro de Biotecnologia
8. Centro de Ciências Biológicas e da Saúde
9. Centro de Ciências e Tecnologias Agropecuárias
10. Centro de Pesquisa Agroflorestal da Amazônia Oriental
11. Centro de Pesquisa Agropecuária de Clima Temperado
12. Centro de Pesquisa Agropecuária dos Cerrados
13. Centro de Pesquisa e Extensao
14. Centro de Pós Graduação e Pesquisa
15. Departamento Agrícola
16. Departamento de Administração e Economia
17. Departamento de Agricultura
18. Departamento de Agronomia
19. Departamento de Biologia
20. Departamento de Biologia Animal
21. Departamento de Ciência da Computação
22. Departamento de Ciência do Solo
23. Departamento de Ciência dos Alimentos
24. Departamento de Ciências Exatas
25. Departamento de Ciências Florestais
26. Departamento de Entomologia
27. Departamento de Fitossanidade
28. Departamento de Medicina Veterinária
29. Departamento de Química e Tecnologia
30. Departamento de Zootecnia

31. Department Of Reproduction Technology
32. Divisão de Ensaios
33. Embrapa Cafe
34. Embrapa Informática Agropecuária
35. Esacma
36. Funedi Fundação Educacional de Divinópolis
37. Instituto de Ciências Agrárias
38. Instituto de Ciências Biológicas
39. Pesquisa
40. Pesquisa e Desenvolvimento
41. Schneider Childrens Hospital
42. Secretaria de Abastecimento e Agricultura de Minas Gerais
43. Serviço de Negócios para Transferência de Tecnologia
44. Setor de Ciências Agrárias
45. Setor de Meio Ambiente
46. Área Técnica

172 unidades diferentes nos currículos de pessoas ligadas à UFLA:

1. Aracaju Sergipe
2. Biotecnologia
3. Campus Fundacional de Lavras
4. Cartografia
5. Centro Nacional de Pesquisa de Solos - Uep Recife
6. Centro Para Desenvolvimento do Talento Cedet
7. Centro Tecnológico Centro Oeste
8. Centro Tecnológico do Sul de Minas
9. Centro Tecnológico do Sul de Minas Ctsm
10. Centro Tecnológico do Sul de Minas Gerais
11. Centro Tecnológico do Triângulo e Alto Paranaíba
12. Centro de Ciências Agrárias
13. Centro de Pesquisa Em Manejo Ecológico de Pragas e Doenças Ecocentro
14. Centro de Pós Graduação e Pesquisa Campus de Divinópolis
15. Cepecafe Setor de Cafeicultura
16. Ciência da Computação
17. Construções Rurais
18. Coordenadoria Geral de Licenciamento Ambiental
19. Cpac
20. Curso de Agronomia
21. Curso de Ciências Biológicas e da Saúde

22. Curso de Engenharia Agrícola
23. Curso de Pós Graduação Em Engenharia Florestal
24. Curso de Pós Graduação Em Estatística e Experimentação Agropecuária
25. Dcf
26. Dcs
27. Departamento de Administração
28. Departamento de Agricultura
29. Departamento de Alimentos
30. Departamento de Antropologia
31. Departamento de Biologia
32. Departamento de Biologia Vegetal
33. Departamento de Bioquímica e Imunologia
34. Departamento de Botânica
35. Departamento de Ciência da Computação
36. Departamento de Ciência do Solo
37. Departamento de Ciências Administrativas
38. Departamento de Ciências Agrárias
39. Departamento de Ciências Biológicas
40. Departamento de Ciências Exatas
41. Departamento de Ciências Florestais
42. Departamento de Clínica e Cirurgia Veterinária
43. Departamento de Defesa Fitossanitária
44. Departamento de Ecologia Geral
45. Departamento de Economia Doméstica
46. Departamento de Economia Rural
47. Departamento de Educação
48. Departamento de Engenharia Agronômica
49. Departamento de Engenharia Agrícola
50. Departamento de Engenharia Agrícola e Ambiental
51. Departamento de Engenharia Florestal
52. Departamento de Engenharia Hidráulica e Sanitária
53. Departamento de Engenharia Química
54. Departamento de Engenharia Rural
55. Departamento de Engenharia de Transportes
56. Departamento de Entomologia
57. Departamento de Entomologia e Fitopatologia
58. Departamento de Estatística e Computação
59. Departamento de Fisiologia
60. Departamento de Fitotecnia
61. Departamento de Hidráulica e Saneamento
62. Departamento de Matemática

63. Departamento de Microbiologia
64. Departamento de Parasitologia Animal
65. Departamento de Produção Animal
66. Departamento de Produção Vegetal
67. Departamento de Produção e Exploração Animal
68. Departamento de Química
69. Departamento de Solos e Engenharia Agrícola
70. Departamento de Tecnologia de Alimentos
71. Departamento de Tecnologia de Alimentos e Medicamentos
72. Departamento de Turismo
73. Departamento de Zootecnia
74. Departamento de Zootecnia de Ruminantes e Animais de Ceco Funcional
75. Department Of Business Administration
76. Divinópolis
77. Divisão de Sensoriamento Remoto
78. Embrapa Arroz e Feijão
79. Embrapa Clima Temperado
80. Embrapa Recursos Genéticos e Biotecnologia
81. Engenharia de Água e Solos
82. Estação de Hidrobiologia e Piscicultura Estação de Piscicultura da Cemig
83. Faculdade de Zootecnia
84. Fazenda Experimental Santa Rita
85. Fertilidade do Solo
86. Fitopatologia Microbiologia
87. Fitotecnia
88. Grupo de Física
89. Laboratorio de Pesquisa do Meio Ambiente
90. Laboratório Central de Biologia Molecular
91. Laboratório Física do Solo
92. Laboratório de Analise de Água
93. Laboratório de Análise de Sementes
94. Laboratório de Análise de Sementes Oficial Supervisor
95. Laboratório de Biologia Molecular de Fungos Filamentosos
96. Laboratório de Bioquímica de Alimentos e Fisiologia Pós Colheita
97. Laboratório de Biotecnologia
98. Laboratório de Ciências Ambientais
99. Laboratório de Controle Microbiológico Comic
100. Laboratório de Cultura de Tecidos e Plantas Medicinais
101. Laboratório de Física do Solo
102. Laboratório de Geoprocessamento

103. Laboratório de Graos e Cereais
104. Laboratório de Grãos e Cereais
105. Laboratório de Microbiologia do Solo
106. Laboratório de Microestrutura e Arquitetura Alimentar
107. Laboratório de Nematologia
108. Laboratório de Proteção de Plantas
109. Laboratório de Recursos Genéticos e Melhoramento Florestal
110. Laboratório de Sementes
111. Laboratório de Tecnologia da Madeira
112. Laboratório de Toxicologia
113. Laboratório de Virologia Vegetal
114. Laboratório de Zootecnia e Nutrição Animal
115. Lavras
116. Lzna
117. Medicina Veterinária
118. Medicina Veterinária Preventiva
119. Microestructura e Estructura Alimentar
120. Nutrição Animal
121. Nutrição de Cães e Gatos
122. Pato Branco
123. Porto Velho
124. Pólo de Tecnologia Em Qualidade do Café
125. Recursos Pesqueiros
126. Setor de Agrometeorologia
127. Setor de Cafeicultura
128. Setor de Cartografia
129. Setor de Cirurgia
130. Setor de Cirurgia Veterinária
131. Setor de Clínica de Grandes Animais
132. Setor de Clínica de Pequenos Animais
133. Setor de Construções Rurais
134. Setor de Construções Rurais e Ambiência
135. Setor de Controle de Poluição
136. Setor de Cínica de Pequenos Animais
137. Setor de Ecologia
138. Setor de Eletricidade e Automação
139. Setor de Engenharia de Água E Solo
140. Setor de Engenharia de Água e Solo
141. Setor de Estatística e Experimentação
142. Setor de Fisiologia
143. Setor de Fisiologia Vegetal
144. Setor de Fisiopatologia da Reprodução

145. Setor de Fitopatologia
146. Setor de Fitotecnia
147. Setor de Fruticultura
148. Setor de Física
149. Setor de Física do Solo
150. Setor de Física e Conservação do Solo e da Água
151. Setor de Genética
152. Setor de Genética e Melhoramento de Plantas
153. Setor de Grandes Culturas
154. Setor de Matemática
155. Setor de Mecanização Agrícola
156. Setor de Mecânica dos Solos
157. Setor de Medicina Veterinária Preventiva
158. Setor de Microbiologia
159. Setor de Microbiologia do Solo
160. Setor de Mineralogia e Química do Solo
161. Setor de Morfologia
162. Setor de Olericultura
163. Setor de Patologia
164. Setor de Patologia Veterinária
165. Setor de Processamento de Produtos Agrícolas
166. Setor de Processamento e Armazenamento de Produtos Agrícolas
167. Setor de Sementes
168. Setor de Virologia Vegetal
169. Seção de Física de Solos
170. Supervisão de Física e Técnicas Especiais
171. Unucet
172. Área de Fitotecnia

7.4 ANEXO D – Áreas e subáreas cadastradas na Plataforma Lattes das pessoas ligadas à UFLA

45 áreas diferentes nos currículos de pessoas ligadas à UFLA:

1. Administração
2. Agronomia
3. Antropologia
4. Biologia Geral
5. Bioquímica
6. Botânica
7. Ciência da Computação
8. Ciência da Informação
9. Ciência e Tecnologia de Alimentos
10. Comunicação
11. Ecologia
12. Economia
13. Educação
14. Engenharia Agrícola
15. Engenharia Civil
16. Engenharia Elétrica
17. Engenharia Mecânica
18. Engenharia Química
19. Engenharia Sanitária
20. Engenharia de Materiais e Metalúrgica
21. Engenharia de Produção
22. Farmácia
23. Filosofia
24. Fisiologia
25. Física
26. Genética
27. Geociências
28. História
29. Imunologia
30. Matemática
31. Medicina
32. Medicina Veterinária
33. Microbiologia
34. Morfologia
35. Nutrição
36. Parasitologia
37. Probabilidade e Estatística

38. Química
39. Recursos Florestais e Engenharia Florestal
40. Recursos Pesqueiros e Engenharia de Pesca
41. Saúde Coletiva
42. Sociologia
43. Turismo
44. Zoologia
45. Zootecnia

157 subáreas diferentes nos currículos de pessoas ligadas à UFLA:

1. Administração Educacional
2. Administração Pública
3. Administração Rural
4. Administração de Empresas
5. Administração de Setores Específicos
6. Agricultura de Precisão de Máquinas Agrícolas
7. Agrometeorologia
8. Agroquímica e Agrobioquímica
9. Antropologia Rural
10. Aqüicultura
11. Bioclimatologia
12. Biogeografia
13. Biologia Molecular
14. Biologia e Fisiologia dos Microorganismos
15. Bioquímica
16. Bioquímica da Nutrição
17. Bioquímica dos Microorganismos
18. Biotecnologia
19. Botânica Aplicada
20. Caracterização Ambiental Utilizando Geoprocessamento
21. Citogenética Vegetal
22. Citogenética animal
23. Citogenética vegetal
24. Citologia e Biologia Celular
25. Ciência de Alimentos
26. Ciência do Solo
27. Ciências Exatas Aplicada à Agricultura
28. Clínica Médica
29. Clínica e Cirurgia Animal
30. Comportamento Animal
31. Conservação

32. Conservação da Natureza
33. Construções Rurais e Ambiência
34. Corrosão Em Meios de Baixa Condutividade
35. Ecologia Aplicada
36. Ecologia Teórica
37. Ecologia Vegetal
38. Ecologia de Comunidades
39. Ecologia de Ecossistemas
40. Ecologia de Populações
41. Economia Internacional
42. Economias Agrária e dos Recursos Naturais
43. Educaçào e Movimentos Sociais no Campo
44. Eletrônica Industrial, Sistemas e Controles Eletrônicos
45. Energia de Biomassa Florestal
46. Engenharia Hidráulica
47. Engenharia Térmica
48. Engenharia de Alimentos
49. Engenharia de Processamento de Produtos Agrícolas
50. Engenharia de Água e Solo
51. Ensino-Aprendizagem
52. Entomologia
53. Entomologia - Controle Biológico
54. Entomologia - Saúde pública
55. Entomologia e Malacologia de Parasitos e Vetores
56. Enzimologia
57. Epidemiologia
58. Epistemologia
59. Estatística
60. Estatística Aplicada à Genética e Ao Melhoramento de Plantas
61. Estatística e Experimentação Agronômica
62. Estruturas
63. Evolução
64. Extensão Rural
65. FISIOLOGIA VEGETAL
66. Farmacognosia
67. Fenômenos de Transporte
68. Fertilidade do Solo e Adubação
69. Fisiologia Geral
70. Fisiologia Vegetal
71. Fisiologia da Reprodução
72. Fisiologia de Sementes
73. Fisiologia de Órgãos e Sistemas

74. Fisiologia do Estresse
75. Fitogeografia
76. Fitopatologia
77. Fitossanidade
78. Fitotecnia
79. Floricultura, Parques e Jardins
80. Formação de Professor
81. Formação de Professores
82. Fundamentos da Educação
83. Física Geral
84. Física Nuclear
85. Física da Matéria Condensada
86. Física das Partículas Elementares e Campos
87. Físico-Química
88. Genética Molecular e de Microorganismos
89. Genética Quantitativa
90. Genética Vegetal
91. Genética e Melhoramento Florestal
92. Genética e Melhoramento Vegetal
93. Genética e Melhoramento de Plantas
94. Genética e Melhoramento dos Animais Domésticos
95. Geofísica
96. Geografia Física
97. Geologia
98. Geometria e Topologia
99. Geoprocessamento
100. Geotécnica
101. Gerência de Produção
102. Hidráulica e Saneamento
103. História Moderna e Contemporânea
104. História do Brasil
105. IRRIGAÇÃO E DRENAGEM
106. Imunologia Celular
107. Informática
108. Inteligência Artificial
109. Irrigação e Drenagem
110. Levantamento dos Recursos Naturais Renováveis
111. Manejo Florestal
112. Matemática Aplicada
113. Matemática da Computação
114. Medicina Veterinária Preventiva
115. Melhoramento de Plantas

116. Metabolismo e Bioenergética
117. Metodologia e Técnicas da Computação
118. Microbiologia
119. Microbiologia Agrícola
120. Microbiologia Aplicada
121. Microbiologia Básica
122. Microbiologia do Solo
123. Morfologia Vegetal
124. Máquinas e Implementos Agrícolas
125. Métodos Quantitativos em Economia
126. Nutrição Animal
127. Nutrição e Alimentação Animal
128. Otimização e Pesquisa Operacional
129. Outras Sociologias Específicas
130. Pastagem e Forragicultura
131. Patologia Animal
132. Plantas Medicinais
133. Probabilidade e Estatística Aplicadas
134. Processos Industriais de Engenharia Química
135. Produção Animal
136. Produção Vegetal
137. Propagação de plantas
138. Qualidade
139. Química Analítica
140. Química Inorgânica
141. Química Orgânica
142. Recursos Genéticos
143. Recursos Hídricos
144. Reprodução Animal
145. Rádio e Televisão
146. Saneamento Ambiental
147. Silvicultura
148. Sistemas Elétricos de Potência
149. Sistemas de Computação
150. Sistemática
151. Sociologia Rural
152. Sociologia do Conhecimento
153. Sociologia do Desenvolvimento
154. Taxonomia dos Grupos Recentes
155. Tecnologia Química
156. Tecnologia de Alimentos
157. Tecnologia e Utilização de Produtos Florestais