

UNIVERSIDADE FEDERAL DE GOIÁS
FACULDADE DE COMUNICAÇÃO E BILIOTECONOMIA
CURSO DE BILIOTECONOMIA

KELLEN NEVES E SILVA

**PERFIL DAS ATIVIDADES DE PESQUISAS E EXTENSÃO REALIZADA NA
UNIVERSIDADE FEDERAL DE GOIÁS: POSSIBILIDADES E LIMITES NA
UTILIZAÇÃO DE MINERAÇÃO DE DADOS**

Goiânia
2008

KELLEN NEVES E SILVA

**PERFIL DAS ATIVIDADES DE PESQUISAS E EXTENSÃO REALIZADA NA
UNIVERSIDADE FEDERAL DE GOIÁS: POSSIBILIDADES E LIMITES NA
UTILIZAÇÃO DE MINERAÇÃO DE DADOS**

Trabalho de conclusão de curso apresentado ao curso de Graduação em Biblioteconomia da Faculdade de Comunicação e Biblioteconomia pertencente a Universidade Federal de Goiás como requisito para obtenção do grau de Bacharel em Biblioteconomia.

Orientador: Prof. Dra. Suely Henrique de Aquino Gomes.

Co-orientador: Prof. Dr. Pedro Henrique Ferri

Goiânia
2008

S623p SILVA, Kellen Neves e.

Perfil das atividades de pesquisa e extensão realizada na Universidade Federal de Goiás: possibilidades e limites na utilização de mineração de dados / Kellen Neves e Silva – Goiânia, 2008.
105 f.: il, enc.

Orientadora: Prof^ª. Dr^ª Suely Henrique de Aquino Gomes.

Co-orientador: Prof^º. Dr^º Pedro Henrique Ferri

Trabalho de conclusão de curso (Graduação) – Faculdade de Comunicação e Biblioteconomia, Universidade Federal de Goiás, 2008.

1. Mineração de dados 2. Gestão do conhecimento 3. Sistemas de gestão do conhecimento. I. Título.

CDU 004.65

KELLEN NEVES E SILVA

**PERFIL DAS ATIVIDADES DE PESQUISAS E EXTENSÃO REALIZADA NA
UNIVERSIDADE FEDERAL DE GOIÁS DO ANO DE 2007: POSSIBILIDADES E
LIMITES NA UTILIZAÇÃO DE MINERAÇÃO DE DADOS**

Trabalho de conclusão defendida no Curso de Graduação em Biblioteconomia da Faculdade de Comunicação e Biblioteconomia da Universidade Federal de Goiás, para a obtenção do título de Bacharel, aprovada em ____ de _____ de 2008, pela Banca Examinadora constituída pelos seguintes professores:

Profº. Drª. Suely Henrique de Aquino Gomes
(orientadora)

Prof. Dr. Pedro Henrique Ferri – Instituto de Química/UFG
(Co-orientador)

Ms. Rogério Arantes Gaioso – IBGE – UE/GOIÁS
(Convidado)

Dedico a Deus, que me deu sabedoria,
saúde durante a realização deste trabalho
Aos amigos da família Lauro e Dóris Andrade
À minha mãe Marta, aos meus irmãos Daniel e Ketlen
À minha tia Maria Ducarmo.

AGRADECIMENTOS

Aos pastores Nilda e Ivan, e Lia que me ajudaram por meio de orações, os quais se alegraram com minhas conquistas.

À minha orientadora Suely Gomes, que acreditou no tema escolhido, pela orientação e dedicação permitindo a realização desse TCC.

Às minhas amigas Denise Moreira e Jaqueline Rocha pelo apoio durante o curso de Biblioteconomia e durante a pesquisa.

Ao CERCOMP que disponibilizaram uma cópia de dados; especialmente ao Fernando, responsável pela divisão de sistema.

Agradecimento em especial ao professor Pedro Ferri, do Instituto de Química pelo grande auxílio na análise dos dados, principal atividade deste trabalho.

Agradeço também a todos os demais professores do Curso de Biblioteconomia, da Faculdade de Comunicação e Biblioteconomia que foram fundamentais na minha formação.

Bem-aventurado o homem que acha sabedoria, e o
homem que adquire conhecimento.

(Provérbios 3:13)

“A sabedoria é a coisa principal;
adquire pois a sabedoria, emprega
tudo o que possues na aquisição de
entendimento”

(Provérbios 4:7)

RESUMO

O objetivo geral deste estudo foi o de verificar as possibilidades e limites da utilização das técnicas de mineração de dados como ferramenta para investigar e analisar as bases de dados gerados pelos sistemas de cadastro de pesquisas e extensão da Universidade Federal de Goiás (UFG). O pressuposto da pesquisa foi a possibilidade de existir uma grande quantidade de informação e conhecimento “escondidos” nos sistemas de cadastro da UFG. A análise dos dados foi realizada através de técnicas como a análise multivariada e análise descritiva dos dados utilizando o *software* SPSS *statistics* 17.0 e o SPAD (Sistema Portátil de Análise de Dados). Conclui-se que a técnica de Mineração de dados, associada a *softwares* especialistas e algoritmos, pode ser um instrumento importante nas Instituições de Ensino Superiores. Na UFG, as ferramentas de mineração de dados poderiam ser utilizadas para a descoberta de padrões e tendências, como uma ferramenta de gestão de competências na tarefa de identificar áreas de conhecimento com escassez de profissionais, e conhecer as linhas de pesquisas dos orientadores e participantes.

Palavras-chave: Gestão do conhecimento. Mineração de dados. Gestão de competências. Sistema de gestão do conhecimento.

LISTA DE ILUSTRAÇÕES

FIGURA 1	Quatro modos de conversão do conhecimento	24
FIGURA 2	Etapas no processo de <i>Data Mining</i> de Fayyad, Piatetsky – Shapiro e Smith	55
FIGURA 3	Metodologia de mineração de dados utilizados para o estudo de caso	67
FIGURA 4	Lista de palavras por ordem de frequência	83
FIGURA 5	Lista de segmentos de palavras por ordem de frequência	85
FIGURA 6	Relação entre as palavras-chaves e as Grandes áreas	86

LISTA DE QUADROS

QUADRO 1	Número das Grandes Áreas na matriz	73
QUADRO 2	Representação das Classes e Grandes Áreas	74

LISTA DE GRÁFICOS

GRÁFICO 1	Quantidade de pesquisas por grandes áreas do conhecimento	76
GRÁFICO 2	Quantidade de pesquisas por áreas do conhecimento	77
GRÁFICO 3	Nível de formação dos orientadores	78
GRÁFICO 4	Pesquisa realizadas por órgãos/ unidades acadêmicas dos orientadores	79
GRÁFICO 5	Tipo de participante	80
GRÁFICO 6	Formação do orientador nas grandes áreas	81
GRÁFICO 7	Tipo de participante por grandes áreas	82

LISTA DE SIGLAS

CAPES	Coordenação de aperfeiçoamento de pessoal de ensino superior
CEDECOM	Centro de Comunicação
CNPq	Conselho Nacional de Desenvolvimento de Pesquisa Científico e Tecnológico
FACOMB	Faculdade de Comunicação e Biblioteconomia
FINEP	Financiadora de estudos e projetos
KDD	<i>Knowledge Discovery in Databases</i> ou Descoberta de conhecimento em Base de Dados
LCC	Laboratório de Computação Científica
MCT	Ministério da Ciência e Tecnologia
PL	Plataforma Lattes
PROEC	Pró-reitoria de extensão e cultura
PRPPG	Pró -reitoria de pesquisa e pós-graduação
RNA	Rede Neural Artificial
SAPP	Sistema de Acompanhamento de Projetos de pesquisa
SICAD	Sistema de cadastro de atividades docentes
UFG	Universidade Federal de Goiás
UFMG	Universidade Federal de Minas Gerais

SUMÁRIO

1	INTRODUÇÃO	16
2	REFERENCIAL TEÓRICO	17
2.1	SOCIEDADE DA INFORMAÇÃO E A ERA DO CONHECIMENTO	17
2.1.1	Dados, informação e conhecimento	18
2.1.2	Tipos de conhecimento	19
2.1.3	Ciclo de conversão de dados em conhecimento	22
2.1.4	Formas de aquisição do conhecimento	24
2.1.4.1	Aquisição de outra organização	25
2.1.4.2	Interação empresa – universidade	25
2.2	A GESTÃO DO CONHECIMENTO	26
2.2.1	Aspectos históricos	26
2.2.2	Aspectos conceituais	27
2.2.3	Gestão do conhecimento e gestão de competências	28
2.2.3.1	Capital humano e capital intelectual	28
2.2.3.2	Gestão de competências: conceitos	28
2.2.4	Importância da gestão do conhecimento	30
2.3	A GESTÃO DO CONHECIMENTO NAS IES	31
2.3.1	Gestão do conhecimento como metodologia de aprendizagem do docente	31
2.3.2	Gestão do conhecimento no contexto das comunidades científicas acadêmicas	32
2.3.2.1	Conceito de grupo de pesquisas	32
2.3.2.2	Perfil de grupo de pesquisas	33
2.3.2.3	Características dos grupos de pesquisas	34
2.3.2.4	Funções do grupo de pesquisa	34
2.3.2.5	Atividades realizadas por grupos de pesquisas	35
2.3.3	Gestão do conhecimento como foco no processo decisório	36
2.4	TECNOLOGIAS DE GESTÃO DO CONHECIMENTO	37
2.4.1	Categorias de sistemas de gestão do conhecimento	37

2.4.2	Sistemas de armazenamento e compartilhamento do conhecimento	38
2.4.3	Iniciativas de construção de sistemas de gestão do conhecimento e competência	39
2.4.3.1	Iniciativas nas IES: Catálogo Quem Sabe da UFMG	39
2.4.3.2	Iniciativas em âmbito nacional: Plataforma Lattes	41
2.5	MINERAÇÃO DE DADOS	45
2.5.1	Conceitos	45
2.5.2	Histórico	47
2.5.3	Etapas de mineração de dados	48
2.5.3.1	Pré-processamento: organização e transformação dos dados	48
2.5.3.2	Mineração de dados: aplicação das técnicas de mineração	51
2.5.3.3	Pós-processamento: interpretação dos resultados e descoberta do conhecimento	51
2.5.4	Tarefas de mineração de dados	53
2.5.4.1	Classificação	53
2.5.4.2	Estimativas	54
2.5.4.3	Predição ou predição	55
2.5.4.4	Análise de agrupamento ou clusterização (segmentação)	55
2.5.4.5	Sumarização	56
2.5.4.6	Descoberta de associação	56
2.5.5	Técnicas de mineração de dados	57
2.5.5.1	Redes Neurais	60
2.5.5.2	Algoritmos Genéticos	58
2.5.5.3	Método estatístico: agrupamento e multivariada.	58
2.5.5.4	Árvores de Decisão	59
2.5.5.5	Regras de associação	59
2.5.6	Componentes de mineração de dados	60
2.5.6.1	Conjunto de dados	60
2.5.6.2	Tecnologia de mineração de dados	61
2.5.6.3	Algoritmos	63
2.5.6.4	Especialista do domínio da aplicação	63
2.5.7	Áreas de aplicação de mineração de dados	64
3	METODOLOGIA	65

3.1	NATUREZA DA PESQUISA	66
3.2	MÉTODOS UTILIZADOS	66
3.2.1	Procedimentos de pesquisa	66
3.2.2	Instrumentos de pesquisa	67
3.3	CONTEXTO DA PESQUISA: ATIVIDADES DE PESQUISAS E EXTENSÃO NA UNIVERSIDADE FEDERAL DE GOIÁS	68
3.3.1	Descrição da instituição	68
3.3.2	Produção científica da UFG	69
3.4	ANÁLISE DOS DADOS.....	69
3.4.1	Identificação dos objetivos	69
3.4.1.1	Análise do Sistema de Cadastro da UFG	70
3.4.2	Obtenção e entendimento dos dados	71
3.4.3	Pré-processamento: metodologia de mineração de dados	72
3.4.4	Mineração de dados	72
3.4.5	Pós - processamento: interpretação dos resultados.....	74
4	RESULTADOS DA PESQUISA	74
4.1	ANÁLISE DESCRITIVA DOS DADOS	75
4.1.1	Quantidade de pesquisas por Grandes áreas do conhecimento	75
4.1.2	Quantidade de pesquisas por Áreas do conhecimento	76
4.1.3	Nível de formação dos orientadores	77
4.1.4	Número de pesquisas existentes por unidade acadêmica	78
4.1.5	Tipo de participante	79
4.2	ANÁLISE MULTIVARIADA DOS DADOS	80
4.2.1	Formação do orientador por Grandes áreas	80
4.2.2	Tipo de participantes por Grandes Áreas	81

4.2.3	Número de palavras-chaves mais freqüentes em relação ao orientador	82
4.2.4	Segmentos de palavras –chaves	83
4.2.5	Relação entre as palavras-chaves e as Grandes áreas	84
4.3	APLICABILIDADE DAS TÉCNICAS DE MINERAÇÃO DE DADOS E MINERAÇÃO DE TEXTOS NA UFG: POSSIBILIDADES E LIMITES	86
4.3.1	Possibilidades da mineração de dados	87
4.3.2	Limites da mineração de dados.....	88
5	CONSIDERAÇÕES FINAIS	90
	REFERÊNCIAS	92
	ANEXOS	
	ANEXO A – Freqüência dos resultados extraído com o software SPSS Statistics 17.0 sobre os dados da pesquisa científica da UFG	99
	ANEXO B - Figuras de formatação dos dados do <i>software</i> SPAD 5.5	104

1 INTRODUÇÃO

Na última década do século XX, as empresas fizeram um grande investimento em tecnologia da informação para melhor coletar, processar, analisar e distribuir as informações.

Na atual Revolução da informação, os negócios procuraram, na informação, as soluções para resolver seus problemas e para reduzir a margem de erro no processo de tomada de decisão. O computador passou a ser o componente principal no processamento de informação, sendo utilizado 24 horas por dia para diferenciar produtos e ampliar os serviços de mercado (WEITZEN, 1991).

Por outro lado, o conhecimento tem sido reconhecido como um dos mais importantes recursos de uma organização, permitindo a tomada de decisões inteligentes nos planos organizacionais e individuais, induzindo a inovação e capacidade de criar produtos e serviços excelentes em termos de complexidade, flexibilidade e criatividade (CARDOSO, 2005).

Neste contexto, as Instituições de Ensino Superiores (IES), são conhecidas como organizações que produzem conhecimento voltado para a produção científica. Estas instituições, tem acumulado grandes bases de dados contendo um grande número de informações sobre a produção científica, englobando não só dados sobre a produção, mas dados sobre pessoas que fazem parte de grupo de pesquisas. De acordo com Cardoso (2005), é preciso transformar esses dados armazenados em informação para que esta seja um instrumento estratégico de apoio à tomada de decisão.

A mineração de dados surge como instrumento importante para a Gestão do Conhecimento, principalmente no apoio a decisão nas organizações privadas e públicas de qualquer ramo de atividade. Um dos objetivos da mineração de dados e de textos é localizar os padrões escondidos em base de dados através do auxílio nas buscas, que podem ser realizadas utilizando ferramentas diversas. Estas ferramentas geram respostas a um conjunto de questões, a partir das filtragens, que por fim geram um relatório com as descobertas feitas para que sejam estudados e interpretados (LIMA, 2006).

O seguinte estudo pressupõe que no banco de dados existente na Universidade Federal de Goiás (UFG) sobre as atividades científicas pode existir conhecimento escondido. Através da mineração de dados, os órgãos teriam acesso às informações importantes que permitiriam uma gestão de atividades científicas e tomada de decisão.

O objetivo geral deste estudo foi verificar as possibilidades e limites da utilização das técnicas de mineração de dados como ferramenta para investigar e analisar as bases de dados dos sistemas de cadastro de pesquisas e extensão da Universidade Federal de Goiás.

Os objetivos específicos foram:

- a) apresentar um estudo teórico e descritivo sobre mineração de dados e gestão do conhecimento;
- b) enumerar as possibilidades e limites das técnicas de mineração de dados;
- c) apresentar um estudo de caso de mineração de textos nos sistemas de cadastro de pesquisa da UFG;
- d) mostrar os resultados obtidos.

Os seguintes problemas que nortearam a pesquisa foram:

- a) A mineração de dados pode ser utilizada como ferramenta de gestão de competências através do levantamento do perfil dos orientadores e atividades de pesquisa e extensão das IES?;
- b) Quais os limites e possibilidades da utilização da mineração de dados como mecanismo de indicadores sobre a produção científica das pesquisas nas IES?.

O estudo é dividido nos seguintes pontos principais: conceitos sobre a Gestão do conhecimento; descrição das principais iniciativas de criação de sistemas especialistas de gestão do conhecimento nas IES; conceitos e características das técnicas de mineração de dados; estudo de caso: análise das bases de dados sobre atividades de pesquisa e extensão da UFG; conclusão abordando as possibilidades e limites da utilização das técnicas de mineração de dados.

2 REFERENCIAL TEÓRICO

2.1 SOCIEDADE DA INFORMAÇÃO E A ERA DO CONHECIMENTO

Ao longo da história mundial é constatado que, desde épocas remotas, os homens que se encontravam na vanguarda do conhecimento destacavam-se dos demais.

Nesta nova economia global, as organizações têm gerido o conhecimento como capital importante para a inovação, sendo conhecido como gestão do capital do conhecimento. Esta nova economia surge através das mudanças que ocorreram nas atividades humanas e nas organizações, com a transformação da antiga economia baseada no lucro para a economia

baseada no conhecimento, tornando-se como um recurso importante para a vantagem competitiva. (DURAND, 1998 apud FLEURY e FLEURY, p.32, 2006).

Assim como ocorreu na década de oitenta, que teve como importância a qualidade da gestão, e na década de noventa, com a importância da reengenharia dos processos, uma evolução natural destas décadas para as organizações é a importância do gerenciamento do conhecimento. (BERNARDES NETO, 2001).

A Era da Informação e do conhecimento surgiu após a Segunda Guerra Mundial, tendo como característica o capital guiado pela informação e pela popularização das novas tecnologias de armazenamento de informações (LOJKINE, 1995). As organizações se despertaram para o novo papel dos indivíduos dentro da organização, onde o capital humano assume um papel de destaque na organização do conhecimento por criar ativos tangíveis e intangíveis, direcionando seus esforços tanto para dentro quanto para fora da empresa. Assim, dentro das organizações que valorizam o conhecimento passa a existir a presença de “funcionários do conhecimento”, com alto grau de qualificação e escolaridade. Com isto, surge a importância da competência na gestão estratégica da empresa (SANTOS; CERANTE, 2000).

A economia da Era do Conhecimento é uma economia baseada não mais em recursos naturais e físicos, como na Era Industrial, mas baseada em recursos como o conhecimento e a comunicação. Enquanto a Revolução Industrial caracteriza-se pelo acúmulo de capitais, fruto da exploração do trabalhador, a Revolução do Conhecimento caracteriza-se pela valorização do conhecimento, que se tornou o principal ativo da empresa enquanto sua administração tornou-se a principal atividade da organização (STEWART, 1997 apud SANTOS; CERANTE, 2000).

2.1.1 Dados, informação e conhecimento

As organizações estão percebendo que não só dados e informação são ativos organizacionais importantes, mas estão reconhecendo o valor do conhecimento.

A gestão do conhecimento depende de uma releitura de todo um contexto que envolve sua natureza, pois envolve elementos como dados, informação e conhecimento. Assim, uma diferenciação destes elementos é importante para entender como funciona o processo de transformação da informação em conhecimento e mostrar a importância desses elementos na sociedade do conhecimento.

Dados, informação e conhecimento são diferenciados como:

- a) **Dados:** Turban; Rainer, Jr.; Potter (2005) conceituam dados como conjunto de letras ou dígitos que descrevem fatos, coisas, eventos, atividades e transações que são registradas, classificadas e armazenadas. Para Davenport (2001), a observação de fatos brutos, ou entidades quantificáveis, pode ser feita por pessoas ou alguma tecnologia;
- b) **Informação:** a informação é definida por como um dado dotado de relevância e propósitos, fazendo parte da primeira fase de processamento dos dados (DRUCKER, [s.d.] apud Davenport, 2001).
- c) **Conhecimento:** Davenport (2001) afirma que o conhecimento é a informação mais valiosa e mais difícil de gerenciar, porque alguém contextualizou de acordo com o seu entendimento, ou sabedoria. A partir de sua contextualização, a informação será utilizada para alguma ação.

Em todos os campos da atividade humana há sempre uma constante busca de elementos de conhecimento como, por exemplo, coletar dados e informes a fim de transformá-los em informações, visando auxiliar a tomada de decisões (VAITSMAN, p. 18, 2001).

2.1.2 Tipos de conhecimento

Ao se falar em conhecimento organizacional é preciso diferenciar dos outros tipos de conhecimento, pois em uma pessoa dentro da organização pode coexistir conhecimento diferente e os tipos de conhecimento podem influenciar na resolução de algum problema ou tomada de decisão.

O conhecimento pode ser divididos em: conhecimento popular, conhecimento teológico, conhecimento filosófico e conhecimento científico.

- a) **Conhecimento popular:** o conhecimento popular é denominado conhecimento vulgar, do senso comum. Tem como origem a experiência do dia a dia; sendo com isto, superficial e falível, subjetivo e psicológico. De acordo com Cervo e Bervian (1996 apud CHERUBINI NETO, 2003), através do conhecimento popular as investigações pessoais são feitas de acordo com as circunstâncias da vida ou então a partir do compartilhamento com outras pessoas, e das tradições da coletividade.

- b) **Conhecimento teológico:** o conhecimento teológico é o conhecimento revelado, aceito pela fé teológica e relativo a Deus. Este conhecimento é apoiado em doutrinas de proposições sagradas e direcionado à compreensão do mundo em sua totalidade (CHERUBINI NETO, 2003);
- c) **Conhecimento filosófico:** de acordo com Barros e Lehfeld (1986 apud CHERUBINI NETO, 2003), a palavra filosofia foi introduzida por Pitágoras, a partir das palavras *Philos* (amigo) e *Sophia* (sabedoria). A filosofia está apoiada basicamente na reflexão, suas hipóteses advêm da experiência e não da experimentação. Esta, por sua vez, fornece um amplo leque de interpretações e, por conseguinte, impressões, opiniões e conclusões diversas. A filosofia evoluiu de certa forma que torna imprescindível o contexto histórico no qual os problemas do homem estão inseridos, e hoje não pode ser reduzida à busca da originalidade conceitual e reflexiva;
- d) **Conhecimento científico:** o conhecimento científico é fundamentado através de métodos que visam revolver problema (s) definido (s), ou uma (s) hipótese (s) a ser (em) confirmada(s), através de processos de pesquisa. São características do conhecimento científico: ser real, contingente, sistemático, verificável, falível e aproximadamente exato (LAKATOS e MARCONI, 1982, p.21 apud CHERUBINI NETO, 2003).

No meio organizacional autores como Nonaka e Takeuchi consideram a existência de dois tipos de conhecimento:

- a) **Conhecimento tácito:** para Turbain, Rainer Jr e Potter (2005), o conhecimento tácito é o acúmulo de aprendizado subjetivo ou experimental, ou seja, o acúmulo de experiências, percepções, conhecimentos, segredos de negócios, conjunto de habilidades, entendimento e aprendizado que uma organização possui. O conhecimento tácito é lento e difícil de transferir, e por ser frequentemente desestruturado, não tem uma forma tangível, é difícil de formalizar ou codificar. Nonaka e Takeuchi (1997 apud BERNARDES NETO, 2001) definem o conhecimento tácito como um conhecimento pessoal interno ao indivíduo, resultado de suas experiências, compartilhamento e troca através de contatos diretos e face a

face com outras pessoas. O conhecimento tácito pode ser comunicado de um modo direto e efetivo, como por contato;

- b) **Conhecimento explícito:** é um conhecimento mais objetivo, racional e técnico, codificado (documentado) em um formato que pode ser distribuído para outros ou transformado em um processo ou estratégia. Em uma organização, compreende as políticas, guias de procedimentos, documentos oficiais, relatórios, projetos, produtos, estratégias, competências centrais da empresa e a infra-estrutura de tecnologia da informação (TURBAIN; RAINER JR; POTTER, 2005). De acordo com Nonaka e Takeuchi (1997 apud BERNARDES NETO, 2001), conhecimento explícito é um conhecimento formal que pode ser empacotado como informação. Pode ser encontrado em documentos de uma organização, como por exemplo: reportagens, artigos, manuais, patentes, pinturas, imagens, vídeo, sons, *software*, etc. Pode, também, ser encontrado nas representações que a empresa tem dela mesma, como: gráficos organizacionais, mapas de processos, padrão de missões, perícias de domínio, etc. A aquisição de conhecimento explícito é indireta; ele deve ser decodificado em um modelo mental, onde é, então, internalizado como conhecimento tácito;

Segundo BORGHOFF e PARESCI (1998 apud BERNARDES NETO, 2001), a informação é convertida em conhecimento através de um processo social e humano de compartilhar entendimentos e através da percepção de como fazer as coisas nos níveis individuais e organizacionais. O conhecimento está localizado nas pessoas que aprendem, transmitem e produzem conhecimentos de maneira cooperativa em sua atividade. Além disso, quando a informação é transmitida de uma pessoa para outra, a pessoa que transmitiu não a está perdendo; e com a utilização desta informação, ela não se apaga (LEVY, 1999 apud SILVA, FRANCISCO e SCANDELARI, 2007).

A conversão dos dados em conhecimento não é feita sem algum processo, ou seja, existe uma série de fatores que envolvem pessoas, o conhecimento existente dentro delas, e o processo de interação.

2.1.3 Ciclo de conversão de dados em conhecimento

O ciclo de conversão de dados em conhecimento tem como principais teóricos Nonaka e Takeuchi, que propõem um modelo conhecido como a Teoria da Criação do Conhecimento nas Organizações.

Nonaka e Takeuchi (1998 apud SANTOS; CERANTE, 2000) estabelecem um caminho para a conversão do conhecimento nas organizações através da interação contínua, que ocorrem ao mesmo tempo entre o conhecimento tácito e o conhecimento explícito, do nível individual até o nível inter organizacional. Afirmam que os conhecimentos tácito e explícito não estão totalmente separados e são mutuamente complementares. Os autores apontam um modelo de conversão do conhecimento mostrando a interação entre o conhecimento tácito e o explícito e as quatro formas distintas que ocorrem entre eles. O modelo de criação do conhecimento proposto por Nonaka e Takeuchi, baseia-se no pressuposto de que o conhecimento humano é criado e expandido através da interação social, e entre o conhecimento tácito e o conhecimento explícito. Os autores apontam as seguintes etapas que fazem parte do ciclo de conversão de dados em conhecimento:

- 1) **conversão do conhecimento tácito em tácito - Geração e Compartilhamento**
→ **Socialização:** a principal meta desta primeira etapa do modelo é estimular a socialização. A interação representa a palavra de ordem nesta etapa de criação do conhecimento, baseando-se no diálogo entre os indivíduos para possibilitar a exposição das suas experiências e habilidades;
- 2) **conversão do conhecimento tácito em explícito - Codificação** → **Externalização:** a codificação do conhecimento tem por finalidade colocar o conhecimento em um formato utilizável para que a empresa possa acessá-lo de maneira adequada. Nesta etapa acontece a externalização do conhecimento tácito da equipe de projeto. Os modelos mentais são expostos e discutidos passando a integrar novos conceitos, que serão transformados em modelos reais para o projeto. Estes modelos, que variam desde protótipos de produtos até mecanismos operacionais que facilitem a prestação de algum serviço na organização;
- 3) **conversão do conhecimento explícito em explícito - Coordenação e Armazenamento** → **Combinação:** esta fase possui o mesmo objetivo da combinação que é o de avaliar os novos conceitos justificados e transformados

em conhecimento explícito para a organização, que em conjunto com os conceitos existentes, serão reunidos de acordo com suas similaridades. O mapeamento e a modelagem das fontes do conhecimento podem ser utilizados, nesta etapa, para a avaliação e categorização do conhecimento quanto sua utilidade e importância. De acordo com Davenport (1998 apud SANTOS; CERANTE, 2000), o mapeamento das fontes, por exemplo, pode ser auxiliado por mapas do conhecimento, que servem como uma espécie de guia para retratar o conhecimento existente na empresa, permitindo a sua localização. Os mapas do conhecimento podem servir como ferramenta para avaliação do estoque do conhecimento na empresa possuindo uma dimensão política que ratifica a importância deste ativo na organização. Uma das iniciativas do armazenamento do conhecimento conhecidas são a criação de bancos de dados do conhecimento, que podem incluir elementos como as páginas amarelas da empresa, instrumento de localização de especialistas da empresa;

- 4) **conversão do conhecimento explícito em tácito - Transferência e Difusão → Internalização:** este é o momento em que os indivíduos, não envolvidos na criação de determinado conhecimento, entram em contato com o mesmo através de sua divulgação, capacitando-os a criticar seus aspectos com base em suas próprias visões e experiências. Esta divulgação pode ser apenas interna na organização ou pode englobar o ambiente externo que participa do negócio.

Neste ciclo, durante a fase de compartilhamento do conhecimento tácito ocorre a “socialização” deste conhecimento. Na criação de conceitos o conhecimento tácito é “externalizado”. Durante a justificação de conceitos a organização decide se estes novos conhecimentos são de seu interesse. A construção do arquétipo é feita por meio da “combinação” do conhecimento explícito. Finalmente, o conhecimento é “internalizado” pelos funcionários da organização para, após, ser novamente “socializado”, dando, assim, início a um novo ciclo de gestão do conhecimento que vai se expandindo em espiral na organização, podendo até mesmo extrapolar as fronteiras desta (CHERUBINI NETO, 2003).

A mineração de dados é um conjunto de técnicas e tecnologias que envolve o tratamento do conhecimento explícito. Com base em Nonaka e Takeuchi, pode-se afirmar que, no processo de construção e consolidação destas tecnologias ocorre anteriormente o

processo de interação entre o conhecimento explícito (dados e informações) e o conhecimento tácito (habilidades e experiências), como mostra a figura abaixo:

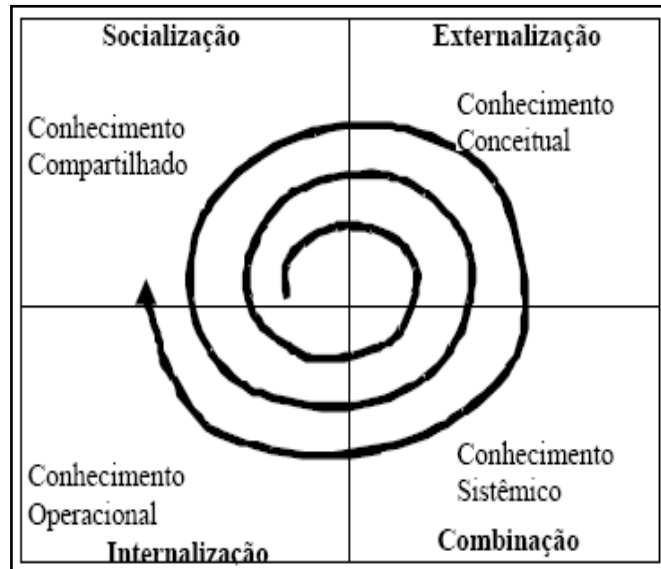


FIGURA 01 - Quatro modos de conversão do conhecimento.
Fonte: NONAKA ; TAKEUCHI (1997 apud CHERUBINI NETO, 2003)

O uso da informação cria significado, constrói conhecimento e facilita a tomada de decisões. A construção do conhecimento leva à inovação, na forma de novos produtos ou novas competências, aperfeiçoa os antigos e melhora seus processos organizacionais (CHOO, 2003, p. 29).

2.1.4 Formas de aquisição do conhecimento

As empresas têm se voltado para o gerenciamento do conhecimento explícito e sistemático a fim de desenvolver o capital intelectual necessário ao sucesso.

Além da utilização da tecnologia para a transformação de dados em conhecimento, de acordo com Santos e Cerantes (2000), o desenvolvimento do conhecimento pode ser feito através, ainda, de duas formas: da aquisição de outra organização e através da interação com as universidades.

2.1.4.1 Aquisição de outra organização

Esta forma ocorre quando uma organização realiza um processo de compra de outra organização ou da contratação de seu pessoal. O sucesso deste processo dependerá do esforço de se procurar manter o ambiente original onde o conhecimento se desenvolveu procurando, inclusive, incluir as pessoas-chaves, oriundas da organização adquirida, além do esforço em localizar e mensurar o valor real do que foi adquirido (SANTOS; CERANTES, 2000).

Como exemplo deste tipo de geração do conhecimento pode-se citar a compra da *Lotus* pela IBM, no ano de 1995, com o intuito não de adquirir os produtos desta primeira empresa, mas principalmente de adquirir o conhecimento de seus funcionários, que possuem capacidade de desenvolver, a exemplo do *Notes*, produtos para uso colaborativo (SANTOS; CERANTES, 2000).

2.1.4.2 Interação Empresa – Universidade

A interação entre empresa e a universidade pode ser realizada através das participações em pesquisas e extensão.¹ Este processo pode envolver certo grau de transferência, mesmo que temporária, cabendo a empresa contratante a responsabilidade por esta aquisição (SANTOS; CERANTES, 2000).

Como exemplo dessa relação entre universidades e empresas privadas está a concessão de apoio financeiro por parte das empresas às instituições universitárias e em troca, por exemplo, do direito de prioridade na utilização dos resultados obtidos pelas pesquisas por elas financiadas. Nos últimos anos, vem crescendo o fomento à pesquisa nas universidades, estimulando a formação de novos pesquisadores e o desenvolvimento de pesquisas nas diversas áreas do conhecimento. Verifica-se, também, que o maior desafio da pesquisa dentro de uma IES está na transferência do conhecimento novo para imediata utilização pela sociedade (LEITE, 2006). Schwartzman (1999 apud LEITE, 2006) ressalta que através das IES os problemas da sociedade e das empresas poderão ser solucionados por meio dos projetos de pesquisa. Assim sendo, uma das questões importantes é a necessidade de estabelecer alianças estratégicas entre universidades e empresas. A interação da universidade

¹ De acordo com Tavares (1997 apud LEITE, 2006), a extensão universitária pode ser compreendida como um processo de comunicação entre a comunidade acadêmica e a sociedade, visando elaborar a prática e aperfeiçoar o conhecimento científico. Esse fluxo que se estabelece permite o confronto entre o conhecimento acadêmico e as diferentes realidades contextualizadas na comunidade. Além disso, a extensão pode ser vista como uma prática acadêmica que visa interligar a universidade em suas atividades de ensino e pesquisa com as demandas da sociedade, buscando respeitar o compromisso social da universidade.

com o mercado de trabalho pode ser vista como uma das ações competitivas para aumentar a “empregabilidade” dos egressos; ao mesmo tempo, constitui-se numa forma efetiva para aumentar e socializar os conhecimentos gerados na universidade, propiciando junto ao mercado inovações que possibilitam a promoção do seu crescimento, provocando o reposicionamento da imagem corporativa da universidade (SANTOS; CERANTES, 2000).

Essa parceria traz vantagem competitiva, tanto para a universidade como para as organizações que compõem o mercado de trabalho. Nesta parceria, a universidade aplica e divulga suas pesquisas, suas novas tecnologias, projetos científicos perante a sociedade e, sobretudo, compreendem as necessidades, anseios e aspirações do mercado a respeito dos serviços gerados (MORAES, 1997 apud LEITE, 2006).

2.2 A GESTÃO DO CONHECIMENTO

2.2.1 Aspectos históricos

O conceito de gestão do recurso do conhecimento, gestão de competências, a gestão do conhecimento ou gestão do capital intelectual surgiu no início da década de noventa, e de lá para cá vem se difundindo nos meios acadêmicos e organizacionais. Entretanto, o interesse pelos assuntos ligados ao conhecimento vem desde a origem da humanidade (NASCIMENTO, 2006).

Dentre os expoentes da gestão de conhecimento são conhecidos autores como: Karl Sveiby com o conceito “trabalhador do conhecimento”; Peter Drucker ; e Nonaka e Takeuchi (com a teoria da criação do conhecimento).

De acordo com Cherubini Neto (2003), desde o seu início até os dias atuais a gestão do conhecimento passou por 3 (três) fases diferentes.

- a) **Primeira fase:** iniciou em 1992 e era voltada para o interior da própria organização, focando a produtividade. Nesta fase, houve um uso intenso da tecnologia da informação como projetos de base de dados, base de dados de melhores práticas, instalação de *Lotus Notes*, etc.
- b) **Segunda fase:** esta fase foi similar à primeira, mas com o foco no cliente de como aproveitar o conhecimento da organização sobre o cliente para melhor atendê-lo. Nesta fase a criação de “*Data-warehousing*” era a vedete.

- c) **Terceira fase:** a fase atual é a terceira, onde a interação é o tema principal. O foco passa a ser a interatividade da tecnologia da informação nas páginas web, *ebusiness*, *e-commerce*, transações *online*, etc.

No entanto, na próxima fase, a quarta, o foco será nas pessoas. A preocupação principal será como maximizar a habilidade das pessoas para criar novos conhecimentos na organização e como construir ambientes que conduzam ao compartilhamento do conhecimento (SVEIBY, 2001 apud CHERUBINI NETO, 2003).

2.2.2 Aspectos conceituais

Para Tarapanoff (2001, apud MORAES, GOMES e FRANKEN, 2005), a gestão do conhecimento é um conjunto de atividades que busca desenvolver e controlar todo tipo de conhecimento em uma organização. O objetivo é a utilização na consecução de seus objetivos e o apoio ao processo decisório. 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.

Turban; Rainer, Jr.; Potter (2005), conceituam a gestão do conhecimento como um processo que ajuda as organizações a identificar, selecionar, organizar, disseminar, transferir e aplicar informações e experiências importantes que fazem parte da memória da organização e que normalmente residem dentro da organização de uma maneira desestruturada.

Terra (2000 apud NASCIMENTO, 2006) defende que a gestão do conhecimento está intrinsecamente ligada à capacidade das empresas em utilizarem e combinarem as várias fontes e tipos de conhecimento organizacional para desenvolverem competências específicas e capacidade inovadora, que se traduzem, permanentemente, em novos produtos, processos, sistemas gerenciais e liderança de mercado.

Cardoso (2005) afirma que a gestão do conhecimento é um conjunto de processos, apoiados por ferramentas de tecnologia da informação, que permite 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) (CARDOSO, 2005).

Assim, dentro de vários conceitos pode-se afirmar que a gestão do conhecimento é um conjunto de processos que, através do auxílio das tecnologias da informação, procura

capturar, armazenar, organizar e disseminar o conhecimento explícito (dados e informações) e o conhecimento tácito (habilidades e experiências).

2.2.3 Gestão de conhecimento e gestão de competências

Na era do conhecimento, tanto as organizações empresariais assim como as IES, o elemento humano tem se tornado uma peça-chave para obter competitividade, produtividade e excelência no mercado.

A gestão do conhecimento e a gestão de competências surgem neste contexto como ferramentas que permitem, ao setor de recursos humanos, gerenciar de forma estratégica o capital humano e intelectual dentro das organizações.

2.2.3.1 Capital humano e capital intelectual

Cavalcanti et al. (2001 apud MORAES, GOMES e FRANKEN, 2005), relaciona o capital humano a um ativo intangível ou conhecimento formal, composta pela capacidade, habilidade e a experiência existentes nas pessoas que integram uma organização. O capital humano é um ativo intangível que pertence ao próprio indivíduo, mas que pode ser utilizado pela empresa para gerar lucro ou aumentar seu prestígio e reconhecimento social.

Capital intelectual é tudo aquilo que uma organização valoriza, está contida nas pessoas, derivada de processos, sistemas, cultura, conhecimentos e habilidades, bem como em seus manuais, normas, valores, bases de dados, tecnologias próprias, marcas, segredos comerciais, entre outros (NASCIMENTO, 2006).

2.2.3.2 Gestão de competências: conceitos

De acordo com Fabro (2002 apud PIKANÇO, 2006), a gestão de competências tem sido adotada por muitas organizações como um modelo de gestão, visando orientar seus esforços para planejar, captar, desenvolver e avaliar, nos diferentes níveis da organização (individual, grupal e organizacional), as competências necessárias à consecução de seus objetivos. A gestão por competências seria responsável pela formação do capital intelectual da organização, maximizando os talentos existentes e em potencial.

O tema competência vem sendo pesquisado desde a década de 1970 e entrou na pauta das discussões acadêmicas e empresariais associado a diferentes formas de compreensão, a saber: as competências do indivíduo, das organizações (as competências essenciais) e ainda, os sistemas de formação de competência (FLEURY; FLEURY, 2006).

O termo competência é utilizado por Sveiby (1998, p.43 apud FREITAS JÚNIOR, 2003), “como sinônimo tanto de saber quanto de conhecimento”. Portanto, é individual e refere-se à competência dos membros da instituição. Relaciona-se àquilo que a instituição é capaz, à motivação dos membros da instituição em promover mudanças, realizar melhorias e criar algo novo. Com a gestão de competências é possível a avaliação e a valorização do capital humano da organização universitária por meio da implantação de uma gestão eficaz de recursos humanos.

Fleury e Fleury (2006) definem competência como: um saber agir responsável e reconhecido, que implica mobilizar, integrar, transferir conhecimentos, recursos, habilidades, que agreguem valor econômico à organização e valor social ao indivíduo.

No mundo empresarial, a palavra competência possui diversos significados; todos ligados às características da pessoa: conhecimento, habilidades e atitudes diante das tarefas, e dos resultados (MCLAGAN, 1997 *apud* FLEURY e FLEURY, 2006).

Entre os profissionais de Recursos Humanos, competência é “um conjunto de conhecimentos, habilidades, atitudes que afetam a maior parte do trabalho de uma pessoa, e que se relacionam com o desempenho no trabalho. A competência pode ser mensurada quando comparada com padrões estabelecidos e desenvolvida por meio de treinamento (FLEURY e FLEURY, 2006).

McClelland (1973 *apud* FLEURY e FLEURY, 2006, p.27) diferencia competência de aptidão, que seria um talento natural da pessoa, o que pode ser aprimorado com habilidade, que seria a demonstração de um talento particular na prática, e de conhecimento, o que a pessoa precisa saber para desempenhar uma tarefa.

No contexto das atividades de pesquisas, a gestão de competências traz as seguintes vantagens: a possibilidade de definir perfis acadêmicos, que favorecerão a produtividade do grupo; o desenvolvimento de equipes orientado pelas competências necessárias às diversas atividades do grupo para o aumento da produtividade e para a maximização de resultados (LEITE, 2006).

A gestão de competências em grupos de pesquisas pode ser realizada através do auxílio de sistemas de gerenciamento do conhecimento como mineração de dados, que poderá

proporcionar o levantamento de perfis e um melhor gerenciamento nas produções de atividades de pesquisas.

2.2.4 Importância da Gestão do conhecimento

A gestão do conhecimento emerge como a área que estuda como as organizações podem entender o que elas se conhecem, o que elas necessitam conhecer e como elas podem tirar o máximo proveito do conhecimento (CARVALHO, 2000 apud CARDOSO, 2005).

Segundo Davenport e Prusak (1998 apud BOTELHO; MONTEIRO e VALLS, 2007), a gestão do conhecimento tem como importância:

- criar repositórios de conhecimento que reúnam tanto conhecimento, quanto informação, principalmente documentos escritos. Estes repositórios podem ser de três tipos: conhecimento externo (inteligência competitiva); conhecimento interno estrutural (relatórios, produtos, procedimentos e técnicas); conhecimento interno tácito ou informal;
- proporcionar ou aumentar o acesso à informação e ao conhecimento, facilitando sua difusão dentro da organização, enfatizando a conectividade, o acesso e a transferência de informação e conhecimento, com base nas tecnologias digitais que garantem a alimentação, o acesso e a transferência em tempo real. O estabelecimento de normas e padrões é essencial para organizar a memória digital;
- criar um ambiente positivo no qual a criação, transferência e uso do conhecimento sejam valorizados. Isso envolve visão, valores e comprometimento das lideranças dispostas a encorajar a criatividade individual e o trabalho em equipes multifuncionais;
- reconhecer o conhecimento como um bem e enfatizar seu valor para a organização.

2.3 A GESTÃO DO CONHECIMENTO NO CONTEXTO DAS IES

De acordo com Cardoso (2005), as IES estão voltadas para a criação e a disseminação do conhecimento científico.

Uma IES tem como matéria-prima o conhecimento e existe para servir a sociedade e contribuir para o seu desenvolvimento, objetivando a formação de profissionais qualificados. As IES estruturam-se por meio de suas atividades acadêmicas e administrativas, objetivando o funcionamento harmônico de suas funções (ensino, pesquisa e extensão) e, ainda, visando assegurar o pleno uso dos seus recursos humanos e materiais (FREITAS JÚNIOR, 2003).

Além de desempenharem seu papel nas atividades de ensino e pesquisa, as universidades vêm fazendo parte do desenvolvimento econômico regional, através de atividades de extensão, cursos, projetos, assessorias e consultorias. Essa participação tem sido organizada permitindo a utilização do conhecimento de docentes e discentes, além da contribuição já realizada por intermédio do ensino, o qual tem como meta a formação de profissionais e no campo da pesquisa, o avanço da ciência e da tecnologia (FREITAS JÚNIOR, 2003).

De acordo com Freitas Junior (2003), a gestão do conhecimento nas IES pode ser estudada a partir de três perspectivas: gestão do conhecimento como metodologia de aprendizagem do docente, gestão do conhecimento no âmbito das comunidades científicas e gestão do conhecimento como foco no processo decisório.

2.3.1 Gestão do conhecimento como metodologia de aprendizagem do docente

De acordo com Freitas Júnior (2003), a primeira abordagem focaliza a gestão do conhecimento como metodologia de aprendizagem nas atividades docentes.

A abordagem que focaliza a gestão do conhecimento como metodologia de aprendizagem nas atividades docentes, considera que o ciclo de vida do conhecimento em uma IES ocorre considerando-se a interação em sala de aula por meio da aplicação de metodologias pedagógicas inovadoras, que favoreçam os processos de socialização e internalização nas atividades docentes (FREITAS JÚNIOR, 2003).

Para isso, é preciso desenvolver um sistema de gestão do conhecimento, baseado em objetos de aprendizagem, que proporcione aos docentes instrumentos para gerenciar e disponibilizar conteúdos e que permite aos alunos construir seu conhecimento por meio do processo de gerenciamento de conteúdos (FREITAS JÚNIOR, 2003).

2.3.2 Gestão do conhecimento no contexto das comunidades científicas acadêmicas

A perspectiva da gestão do conhecimento científico no contexto das comunidades científicas, está relacionada com a gestão do conhecimento produzido por uma disciplina, pesquisa científica, tópico ou campo específico do conhecimento (LEITE, 2006).

Para Leite (2006), a gestão do conhecimento científico no âmbito de comunidades acadêmicas não possui um caráter institucional, e sim disciplinar. Nesse sentido, existe uma caracterização do conceito de comunidade científica como um conjunto de pesquisadores que compartilham interesse sobre tópicos ou áreas específicas, em nível internacional, vindo ao encontro daquilo que caracteriza as redes.

As comunidades acadêmicas são agrupamentos de membros de uma instituição acadêmica envolvidos com atividades de ensino e pesquisa, compartilhando ou não interesses comuns em seus tópicos de estudo. Porém, os pesquisadores pertencem, individualmente, a grupos de interesse em tópicos específicos, sem limites geográficos denominados comunidades científicas (LEITE, 2006).

Nessas comunidades científicas há um constante e intenso compartilhamento de conhecimento, constituindo, no seu tópico específico, o principal *locus* de comunicação e troca de conhecimentos. Todas as atividades de pesquisa envolvem um elevado grau de compartilhamento de conhecimento, tanto o conhecimento tácito quanto o conhecimento explícito. A produção do conhecimento científico não é um processo individual, mas um resultado da colaboração e interação entre cientistas (LEITE, 2006).

2.3.2.1 Conceito de grupo de pesquisas

Grupo de pesquisa pode ser conceituado em três aspectos: como um conjunto de indivíduos organizados em hierarquia, como células com interesses comuns e como microunidades de comunidades científicas.

Segundo o CNPq (Conselho Nacional de Pesquisa), um grupo de pesquisa pode ser entendido como um conjunto de indivíduos organizados hierarquicamente em torno de uma ou, eventualmente, duas lideranças. Essa hierarquia tem como fundamento organizador a experiência e competência no campo científico ou tecnológico, existindo envolvimento profissional e permanente com atividades de pesquisa e que, em algum grau, compartilham recursos (LEITE, 2006).

Leite (2006) considera grupos de pesquisa como células onde um conjunto de pesquisadores se organiza em torno de um determinado tópico, pelo qual nutrem interesse comum. Os grupos de pesquisa são formados por pesquisadores experientes, alunos de doutorado e mestrado e alunos de graduação envolvidos em projetos de pesquisa, ambiente no qual há um intenso e constante compartilhamento de conhecimento científico. Normalmente, as atividades científicas em uma universidade estão vinculadas a esses grupos de pesquisa (LEITE, 2006).

Embora sejam formalmente organizados, os grupos de pesquisa formam um agrupamento de pesquisadores baseado em interações sociais e compartilhamento informal de recursos. Podem ser vistos como elementos ou microunidades de comunidades científicas. No seio dessa informalidade muito conhecimento é criado e compartilhado, especialmente o conhecimento científico tácito (LEITE, 2006).

2.3.2.2 Perfil de grupo de pesquisas

Os grupos de pesquisa estão localizados em universidades, instituições isoladas de ensino superior, institutos de pesquisa científica, institutos tecnológicos, laboratórios de pesquisa e desenvolvimento de empresas estatais ou ex-estatais e em algumas organizações não-governamentais com atuação em pesquisa (ROMÃO, 2002 apud FREITAS JÚNIOR, 2003).

Os pesquisadores e alunos de mestrado e doutorado das universidades brasileiras, por meio dos programas de pós-graduação, são os responsáveis pela produção do conhecimento científico e tecnológico nacional, e constituem os protagonistas do processo de criação do conhecimento científico (LEITE, 2006).

Dentre as atividades que mais intensamente estão relacionadas à criação de novos conhecimentos científicos, pode-se relacionar: pesquisas científicas realizadas por pesquisadores docentes, atividades e pesquisas realizadas por grupos de pesquisa, orientação e desenvolvimento de teses e dissertações (LEITE, 2006).

2.3.2.3 Características dos grupos de pesquisas

Os grupos de pesquisa costumam ser construídos considerando as informações presentes na base de dados do Diretório de Grupos de Pesquisa no Brasil, cujo objetivo é oferecer um suporte atualizado das atividades de pesquisa (FREITAS JÚNIOR, 2003).

Um grupo de pesquisa é formado pelo líder e pesquisadores. Um grupo pode ter no máximo dois líderes. O líder de grupo é a pessoa que detém a liderança acadêmica e intelectual no seu ambiente de pesquisa. Normalmente, possui a responsabilidade de coordenação e planejamento das tarefas realizadas pelo grupo. Sua função consiste em aglutinar os esforços dos demais pesquisadores e apontar diretrizes e novas áreas de atuação dos trabalhos desenvolvidos no grupo. Cabe também ao líder a definição de quais são os pesquisadores, estudantes e técnicos do grupo. No caso de estudantes, é necessário que estejam matriculados em um curso de pós-graduação (mestrado ou doutorado) (FREITAS JÚNIOR, 2003).

Dessa forma, os grupos de pesquisa possuem características próprias. Sua dinâmica é determinada por normas e expectativas de seus pesquisadores e pelos mecanismos de apoio de que dispõem, os quais, se bem definidos e gerenciados, possibilitarão direcionar esforços na busca do alto desempenho de seus membros (FREITAS JÚNIOR, 2003).

2.3.2.4 Funções do grupo de pesquisa

Um grupo de pesquisa necessita satisfazer a duas funções básicas: funções de tarefa e funções de manutenção.

As funções de tarefa correspondem aos atos e comportamentos relacionados à produtividade e dirigidos à realização das atividades do grupo. Essas funções refletem o aspecto racional e operacional das atividades a serem realizadas pelo grupo para o atendimento de metas e objetivos (BOWDITCH; BUONO, 1992 apud FREITAS JÚNIOR, 2003).

As funções de manutenção referem-se aos atos e comportamentos que exprimem as necessidades emocionais e sociais dos membros do grupo. A ênfase está em satisfazer essas necessidades, bem como em estabelecer e manter relações interpessoais de cooperação, coesão e uma atitude favorável em relação ao grupo. Um grupo eficaz é aquele que apresenta desempenho na tarefa e manutenção dos recursos humanos (BOWDITCH; BUONO, 1992 apud FREITAS JÚNIOR, 2003).

2.3.2.5 Atividades realizadas por grupos de pesquisas

De acordo com Lichtnow (2001 apud FREITAS JÚNIOR, 2003), as atividades desenvolvidas em um grupo de pesquisa envolvem um conhecimento altamente especializado, além de facilidade de acesso às novas informações. Entre as principais atividades citadas pelo autor que são realizadas dentro de um grupo de pesquisa e que podem ser suportadas por meio do conhecimento, experiência e habilidade, estão:

- **cooperação de pesquisa e/ou elaboração de projetos de pesquisa:** essa atividade requer conhecimento sobre programas de pesquisa e, também, conhecimentos que auxiliem no planejamento e execução de projetos de pesquisa;
- **desenvolvimentos de modelos conceituais e construção de protótipos:** essa atividade requer o desenvolvimento de modelos conceituais para posterior desenvolvimento de um protótipo. Assim, alguns projetos de pesquisa demandam a construção de protótipos, visando ajustar as funcionalidades das soluções implementadas, antecipando problemas potenciais;
- **elaboração e publicação de trabalhos científicos:** essa atividade consiste na elaboração de trabalhos científicos que instrumentalizam as ações de ensino e pesquisa ou que são resultantes delas. Inclui relatórios técnicos, artigos, dissertações e teses; participação em cursos e eventos. Essa atividade consiste em manter informações atualizadas sobre cursos e eventos de interesse do grupo (conferências, simpósios, congressos, *workshops* etc.) a serem realizados; pesquisa e revisão da literatura.
- **aprendizagem:** essa atividade enfatiza a necessidade da utilização de tutoriais e de material relacionado às áreas de atuação do grupo que possa favorecer a essa atividade;
- **desenvolvimento de novos projetos de pesquisa:** essa atividade requer conhecimento relativo às características dos projetos de pesquisa e aos requisitos envolvidos no seu desenvolvimento e execução.

As atividades desenvolvidas nos grupos de pesquisa requerem conhecimentos especializados e facilidades de acesso a novas informações, conhecimentos e tecnologias. Essa premissa apresenta-se cada vez mais verdadeira à medida que os referidos grupos

tornam-se multidisciplinares e com seus pesquisadores atuando em rede. A análise dessas atividades enfatiza também a necessidade de que o conhecimento necessário à realização dessas seja gerenciado de forma sistemática dentro dos grupos de pesquisa, além de demonstrar claramente os benefícios que tal gerência pode trazer para a realização das atividades (FREITAS JUNIOR, 2003).

2.3.3 Gestão do conhecimento como foco no processo decisório

Para Davenport (1998 apud FREITAS JÚNIOR, 2003), um processo é um conjunto de atividades que consome recursos e produz bens ou serviços.

De acordo com Freitas Júnior (2003), os processos de uma IES podem ser considerados como processos produtivos (envolvidos nas funções de ensino, pesquisa e extensão) ou de apoio (voltados para o atendimento dos órgãos internos da instituição). Os processos produtivos são aqueles diretamente envolvidos nas funções de ensino, e as atividades-fim concentram todas as atividades relativas ao ensino, à pesquisa e à extensão que constituem os objetivos primordiais para os quais se criou a universidade.

Para Freitas Júnior (2003), esta abordagem focaliza a gestão do conhecimento como auxílio ao processo decisório, visando tornar a instituição mais inteligente e competitiva, com vistas à obtenção de melhores resultados possíveis em termos de produtividade e capacidade de inovação. Esta abordagem tem como seguintes objetivos:

- criar uma memória organizacional proveniente do ambiente interno e externo da instituição;
- auxiliar o processo decisório nos diversos setores da instituição;
- propiciar a melhoria do acesso ao conhecimento por meio do uso de tecnologias colaborativas;
- criar e manter perfis de competências dos membros da instituição.

Dentro deste contexto surge a gestão de competências que visa criar uma memória organizacional contendo perfis de competências dos membros da instituição.

2.4 TECNOLOGIAS DE GESTÃO DO CONHECIMENTO

A tecnologia da informação vem sendo utilizada pelas organizações como um recurso a ser aplicado na coleta, armazenamento, produção e apresentações de informações confiáveis e relevantes.

Terra (2000, apud SILVA, FRANCISCO e SCANDELARI, 2007) afirma que a associação entre tecnologia da informação e gestão do conhecimento está relacionada ao uso de sistemas de informação para o compartilhamento de informações ou conhecimento.

De acordo com Rezende (2006), os sistemas de conhecimento podem ser trabalhados juntos com os sistemas de informação, a tecnologia da informação e o capital intelectual, contribuindo para a geração de informações oportunas, através da disponibilização e a gestão do conhecimento de forma personalizada.

2.4.1 Categorias de sistemas de gestão do conhecimento

Santos e Cerantes (2000) dividem os sistemas de gestão do conhecimento nas seguintes categorias:

- a) **Destinadas ao armazenamento do conhecimento - *Knowledge Repository***: Um sistema de armazenamento e gerenciamento de conhecimento envolve ferramentas que permitem a gestão do conhecimento com base em competências. Estes sistemas são conhecidos como mapas do conhecimento, páginas amarelas e diretório de especialistas, que são utilizados para o mapeamento de competências. Estas tecnologias estão relacionadas com o conhecimento explícito, e estão inseridas na classificação, isto é, a organização sabe que tipo de conhecimento possui e onde se localiza (“*You know what you have*”), restando gerenciá-lo, integrá-lo e organizá-lo, de forma a poder tirar o maior proveito de sua extração (SANTOS; CERANTES, 2000);
- b) **Destinadas ao compartilhamento do conhecimento - *Knowledge Sharing***: são grupos de tecnologia em que estão inseridas todas aquelas ferramentas que visam dar suporte ou contribuições para o compartilhamento de informações e a integração entre os funcionários na organização, priorizando o conhecimento tácito. As principais ferramentas estratégicas conhecidas deste grupo são as de *Groupware* (tendo como

principal representante o *Lotus Notes*) e as *Intranets / Internets*. Através destas ferramentas poderá haver uma interação entre a organização e o conhecimento tácito, que esta sabe existir mas não sabe como captá-lo (“*You know what you don't have*”) (SANTOS; CERANTES, 2000);

- c) **Destinadas ao descobrimento do conhecimento - *Knowledge Discovery***: são grupos onde estão inseridas as ferramentas que auxiliam no processo de localização de novos conhecimentos, e a organização, apesar de possuir, não conseguiu identificar (“*You don't know what you have*”), sendo inseridas desta forma na categoria. Como exemplo destas ferramentas, pode-se citar *Data Mining* e *Text Mining*, que permitem trabalhar, através de algoritmos mais poderosos envolvendo inteligência artificial, com os dados estruturados (no primeiro caso) e não estruturados (no segundo caso) para extrair e explorar novos conhecimentos a fim de enriquecer a criatividade das “Equipes do Conhecimento” (SANTOS; CERANTES, 2000);

2.4.2 Sistemas de armazenamento e compartilhamento do conhecimento

Os sistemas de gerenciamento de conhecimento visam armazenar dados sobre as competências de profissionais, a fim de permitir a seleção dos que satisfazem a certas combinações de competências requeridas na formação de equipes de projeto, no preenchimento de cargos, no planejamento de treinamento e outras aplicações. O objetivo é identificar e mapear “quem sabe o que” dentro da empresa (SETZER, 2001).

Neste banco de dados, o conhecimento é codificado e estocado e depois disponibilizado para todos os membros da organização (sendo relevante para o conhecimento explícito), por meio do indivíduo, que o disponibiliza para os demais membros por sua rede de interações (sendo relevante para o conhecimento tácito) (HANSEN, NOHRIA e TIERNEY, 1999 apud FLEURY e FLEURY, p.45, 2006).

A consulta das informações seria possibilitada através de um sistema de busca. No sistema de busca o usuário digita um conjunto de palavras-chave ou áreas de conhecimento e o sistema retornaria não documentos, mas sim pessoas. Além disso, através de ferramentas de mineração de dados, o sistema classificaria os especialistas mais indicados que poderiam colaborar com algum projeto, através do acesso ao currículo das pessoas indicadas pelo sistema. Um recurso como esse cria oportunidades fantásticas de compartilhamento de conhecimento tácito (BARONI, 2005).

Dois exemplos de sistemas de armazenamento e compartilhamento do conhecimento que serão abordados neste estudo são: a Plataforma *Lattes* e o Catálogo de especialistas Quem Sabe da UFMG.

Uma ferramenta bem conhecida, que pode conter dados sobre competências e especialidades de docentes, pesquisadores e profissionais, é a Plataforma *Lattes*, mas no meio acadêmico têm existido diversas iniciativas que visam auxiliar no processo de gestão do capital humano.

Algumas instituições têm construído ferramentas que permitem conhecer, divulgar o perfil e competências dos docentes e pesquisadores e o que eles produzem, e também para conhecer especialistas necessários para dar suporte a um determinado processo. Basicamente, tais sistemas funcionam de maneira similar a um mecanismo de busca (CONNECTT, 2006).

O sistema de descoberta do conhecimento será tratado a parte englobando a mineração de dados.

2.4.3 Iniciativas de construção do sistema de mapeamento de competência e conhecimento

2.4.3.1 Iniciativas nas IES: Catálogo Quem Sabe da UFMG

Dentre as iniciativas existentes nas IES, no desenvolvimento de ferramentas de gestão de pessoas e competências, é conhecido o cadastro de especialistas da Universidade Federal de Minas Gerais (UFMG).

Um banco de dados de especialistas é de vital importância em organizações como laboratórios de ensino, pesquisa e desenvolvimento, inseridos em Instituições de Ensino Superior (STRAUHS, ABREU e RENAUX, 2001).

O Quem Sabe é uma ferramenta de busca desenvolvida pela UFMG com a função de ser um Catálogo de Especialistas e Especialidades da universidade, ou seja, uma ferramenta de busca que indicam especialistas e suas especialidades.

A ferramenta foi desenvolvida pelo Centro de Comunicação (CEDECOM) e pelo Laboratório de Computação Científica (LCC) para facilitar o acesso direto a fontes de informação e criar a possibilidade de visualizar, de forma rápida, o universo de questões com as quais a Universidade trabalha (CAMISASCA, 2005).

Além disso, permite democratizar o acesso da imprensa e de outros segmentos sociais ao conhecimento produzido na Universidade e aos produtores desse conhecimento.

Finalidade

Inicialmente, a finalidade do Quem Sabe era de ser um guia de fontes para os profissionais de imprensa. Mas a idéia ganhou uma dimensão muito maior, tendo sido concebido para ser útil tanto para os meios de comunicação quanto para as empresas e instituições, que necessitem de consultoria, prestação de serviço ou até mesmo de uma informação mais especializada (CAMISASCA, 2005).

A tarefa de produzir o Quem Sabe foi facilitada pela existência de um diretório corporativo, base que concentra informações sobre pessoas e seus vínculos com uma organização. Na UFMG, esse diretório é a espinha dorsal do projeto Grude. O catálogo é um exemplo dos benefícios que uma estrutura consolidada como o Grude pode proporcionar (CAMISASCA, 2005).

Os integrantes do Quem Sabe são os professores da UFMG cujos contratos estejam ativos, excluindo-se os professores aposentados ou exonerados (CAMISASCA, 2005).

Essa ferramenta é muito interessante tanto para o público externo quanto para o público interno, que terão a oportunidade de conhecer melhor a Universidade e as possibilidades que ela oferece (CAMISASCA, 2005).

Estrutura

A busca no Quem Sabe pode ser feita através de busca por especialista, por especialidades e por termo livre. As principais estruturas de busca do sistema são:

- a) **Especialista** ou **Especialidade**: essas opções são mais adequadas quando o usuário tem conhecimento do nome de um Especialista ou sabe o nome de uma especialidade já cadastrada;
- b) **Busca** (dentro do *link* Especialista e especialidade): na opção “Busca” do menu à esquerda, basta selecionar a opção “Especialista” ou “Especialidade” e clicar no nome do professor ou da especialidade. O resultado exibirá somente os nomes das “Especialidades” ou “Especialistas” já cadastrados no banco de dados;
- c) **Termo livre**: o usuário pode realizar a procura através de palavras quaisquer. Nesse caso, o resultado aponta para diversas áreas de conhecimento e o usuário escolhe a

melhor alternativa para sua busca. Em todas as buscas pode-se usar mais de uma palavra que podem vir separadas por espaço, vírgula ou ponto e vírgula.

A página do Catálogo Quem Sabe , ainda tem como recurso de busca as seguintes alternativas:

- a) O *Link* ‘**Todas as especialidades**’ exibe uma lista contendo todas as especialidades cadastradas no sistema independente de possuírem ou não especialistas associados.
- b) O *Link* ‘**Todos os professores**’ mostra uma lista contendo todos os professores da UFMG, independente de possuírem ou não alguma especialidade associada. Este *link* retornará somente professores cadastrados como especialistas.

O Quem Sabe da UFMG proporciona ao usuário 2 (dois) meios de consulta: a Consulta às chamadas “árvores”, que é o próprio banco de dados, onde estão listadas as Especialidades e os Especialistas; e a opção “Busca” Os termos de busca das especialidades foram organizados pela jornalista Ana Maria Vieira. Os termos é uma adaptação da árvore do conhecimento do CNPq, com elementos construídos na UFMG. As categorias estão classificadas em 3 (três) níveis (Área, Sub-área, e Sub-sub-área), buscando facilitar a busca (CAMISASCA, 2005). (CAMISASCA, 2005).

Para conferir a compatibilidade dos Quem Sabe quanto a navegação da internet, foram feitos testes no Mozilla 1.4 (Windows) e Internet Explorer (versão 5). Foi verificado que o sistema mostra-se mais compatível com o Internet Explorer, em qualquer versão.

2.4.3.2 Iniciativas em âmbito nacional: Plataforma Lattes (PL)

No contexto da Gestão do conhecimento, a Plataforma Lattes (PL) apresenta-se como um importante instrumento para a geração do conhecimento, quanto utilizado para a mineração de dados.

Na mineração de dados em âmbito das IES, a PL pode ser utilizado para enriquecer informações sobre os especialistas da universidade ou grupos de pesquisas, para ser incorporado no banco de dados existentes ou para fins de comparação.

Objetivos da PL

Lançada em 16 de Agosto de 1999, a PL é um conjunto de sistema de informação, bases de dados e portais de internet, concebida para integrar os sistemas de informação das agências federais, racionalizando o processo de gestão de ciência e tecnologia. A construção da PL permitiu o início do processo de compatibilização dos sistemas computacionais das agências governamentais: Conselho Nacional de Desenvolvimento de Pesquisa Científico e Tecnológico (CNPq), Financiadora de estudos e projetos (FINEP), Ministério da Ciência e Tecnologia (MCT) e a Coordenação de aperfeiçoamento de pessoal de ensino superior (CAPES) (CARDOSO, 2005).

O objetivo da PL é a coleta integrada e a disponibilização de informações junto aos seus usuários, para o apoio das atividades de planejamento, avaliação e acompanhamento ao fomento do CNPq. A partir do Currículo Lattes, o CNPq desenvolveu um formato-padrão para coleta de informações curriculares, hoje adotados não só pela Agência, mas também pela maioria das instituições de fomento, universidades e institutos de pesquisa do País (CARDOSO, 2005).

A PL visa dar transparência às ações relacionadas à função do CNPq, disponibilizando informações sobre o histórico de bolsistas e de pesquisadores beneficiados nos 10 (dez) últimos anos, por área do conhecimento, instituição, unidade da federação e modalidade da bolsa ou auxílio (CARDOSO, 2005).

No CNPq suas informações são aplicadas em diversas atividades:

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

Tipo de informações oferecidas

No ano de 2008, a PL disponibiliza em sua base de dados cerca de 1.140.000 (um milhão, cento e quarenta mil) currículos de pesquisadores, professores, alunos e vários profissionais das mais diversas áreas do conhecimento que atuam em ciência, tecnologia e inovação. Fazem uso desse sistema: pesquisadores, estudantes, gestores, profissionais e demais atores do Sistema Nacional de Ciência, Tecnologia e Inovação (LATTES, 2008).

A PL permite as organizações empresariais e as IES reunirem dados confiáveis para a análise de C&T. A bases de dados da PL permitem apresentar um perfil de C&T mostrando indicadores importantes quanto a: distribuição por áreas do conhecimento (pesquisadores, estudantes, pós-graduação, produção C&T); e distribuição demográfica dos grupos de pesquisa, pesquisadores e avaliação dos cursos de pós-graduação por região do país (GONÇALVES, 2000).

Estrutura da Plataforma Lattes

De acordo com Cardoso (2005), a Plataforma Lattes integra, atualmente, 4 (quatro) plataformas:

- a) **Sistema Eletrônico de Currículos (*Currículo Lattes*)**: se refere a um sistema eletrônico que registra a vida pregressa e atual dos pesquisadores;
- b) **Diretório dos Grupos de Pesquisa no Brasil**: uma base de dados que registra todos os grupos de pesquisa em atividade no país;
- c) **Diretório de Instituições**: instituições estas que demandam fomento ao CNPq;
- d) **Sistema Gerencial de Fomento**: possibilita uma gestão estratégica para dar mais qualidade às atividades de fomento do CNPq.

Esses 4 (quatro) sistemas de informação são integrados, articulados com outras bases de dados, localizadas fora da agência, como a base de patentes do Instituto Nacional de Propriedade Industrial (INPI), os bancos de dissertações e teses das universidades, constituindo-se na Plataforma Lattes (CARDOSO, 2005).

De acordo com Silva (2007), a forma de cadastro e preenchimento dos campos da PL foi modificada ao longo dos anos e continua em processo de aperfeiçoamento.

Atualmente, a PL é dividida em 8 (oito) módulos, e cada módulo possui uma interface de preenchimento disponível:

- a) **Dados gerais:** Apresenta campos para o registro de informações básicas quanto aos dados pessoais do usuário;
- b) **Projetos:** O usuário informa os dados gerais da pesquisa, características, unidade responsável, equipe, etc. ;
- c) **Produção bibliográfica:** Esse módulo agrupa informações a respeito da produção bibliográfica realizada pelo usuário;
- d) **Produção técnica:** Esse módulo agrupa informações a respeito da produção técnica realizada pelo usuário, como projetos, processos de transformação, realização de curso, etc.;
- e) **Orientações:** O módulo agrupa informações a respeito de orientações ou supervisões realizadas pelo usuário;
- f) **Produção cultural:** O módulo Produção cultural agrupa informações a respeito da produção artística ou cultural do usuário;
- g) **Evento:** O módulo Evento agrupa informações a respeito de eventos que o usuário tenha participado (organização, bancas julgadoras, e etc);
- h) **Banca:** O módulo Banca agrupa informações a respeito das bancas nas quais o usuário esteve presente.

O currículo tem estrutura hierárquica. Cada módulo contém várias áreas e dentro de cada área existe um texto de ajuda que indica como a área deve ser preenchida.

Além dos módulos, a Plataforma *Lattes* também oferece outros recursos. Estes recursos se localizam na página inicial do sistema, apresentando-se como ícones. Estes ícones permitem aos usuários outras funcionalidades como: Dicionário de autores (controle de autores que fizeram parte dos trabalhos realizados pelo usuário); exportação de produções (importações de outros currículos); dicionário de palavras-chave (permite o controle de palavras-chave por parte do usuário; dicionário de Áreas do Conhecimento); dicionário de Instituições; indicadores de produção; visualização de currículo em XML/RTF; impressão e saída do sistema.

A Plataforma conta também com o recurso de busca que permite o acesso além da versão em inglês para a entrada e a recuperação dos currículos. O motor de buscas contém filtros para acesso mais rápido às informações, mesmo nas buscas mais extensas e com palavras-chave altamente citadas nos currículos. A busca passa, ainda, a exibir a frequência com que as palavras-chave informadas aparecem nos currículos recuperados, com ordenação por campo (LATTES, 2008).

Características da Plataforma Lattes (PL)

De acordo com Silva (2007), a Plataforma *Lattes* é um sistema aberto.² Numa terminologia da Ciência da Informação, isso implica dizer que se trata de um sistema com um baixo nível de controle. O controle se diz respeito às representações lingüísticas que são usadas no preenchimento dos currículos, excluindo-se, portanto, questões de segurança no acesso ou rastreamento do comportamento/ interação dos usuários com o sistema.

Segundo Silva (2007), o cadastro de especialistas tem as seguintes características:

- a) **Autonomia total:** o usuário pode cadastrar as palavras que desejar, sem restrição ou qualquer direcionamento. Exemplos: nome de autores, identificação, endereço;
- b) **Autonomia parcial:** são campos que tem autonomia total, mas cada novo termo cadastrado pelo usuário é automaticamente armazenado no sistema, que vai criando uma lista de termos exclusiva do usuário. O usuário pode excluir qualquer termo dessa lista. Exemplos: palavras-chave, nome de autores;
- c) **Sem autonomia:** o sistema prevê opções que o usuário deve selecionar, embora essas opções prévias não impedem que novos termos sejam incluídos. Exemplo: idioma de publicação, áreas do conhecimento, setores de aplicação.

2.5 MINERAÇÃO DE DADOS

Este tópico abordará sobre Mineração de dados, descrevendo suas características, etapas, tarefas e técnicas.

2.5.1 Conceitos

Na literatura sobre *Data mining*, podem ser encontradas algumas definições importantes sobre esta ferramenta.

² A Teoria Geral dos Sistemas define que os sistemas abertos interagem mais com o meio, adaptando-se às mudanças em busca da própria sobrevivência, mantendo contínuas interações com o ambiente que o envolvem. Em sistemas abertos a vulnerabilidade decorre do baixo nível de controle da situação (BERTALANFFY, 1997 apud SILVA, 2007).

Quoniam (2001) afirma que *Data Mining* são todas as técnicas que permitem extrair conhecimento de uma massa de dados que, de outra maneira, permaneceria escondidos.

De acordo com Carvalho (2005), o *Data Mining* pode ser definido como o uso das técnicas automáticas de exploração de grandes quantidades de dados de forma a descobrir novos padrões e relações que, devido ao volume de dados, não seriam facilmente descobertas a olho nu pelo ser humano.

Para Bretzke (2000), o *Data Mining* compreende um conjunto de técnicas e critérios de avaliação qualitativa e quantitativa, modelos de análise, formulação, implementação, etc., diretamente ligados à tecnologia de banco de dados e aos sistemas de suporte à decisão. Este tipo de ferramenta pesquisa grandes bases de dados, procurando por padrões de comportamento e respostas que permitam prever, com a maior acuidade possível, o comportamento de clientes.

Para Jambu (2000 apud PENTEADO 2006), *Data Mining* é o processo que combina vários métodos matemáticos, estatísticos ou combinação de algoritmos, para determinar uma solução para um problema, em um universo decisional.

O *Data Mining* é uma técnica que só se aplica a grandes massas de dados, pois necessita de uma grande quantidade de informações armazenadas para calibrar seus algoritmos e extrair, dos dados, conclusões confiáveis. O *Data Mining* também é considerado um sistema especialista, ou seja, sistemas que imitam o raciocínio humano em um certo ramo do conhecimento para responder a perguntas e sugerir ações (CARVALHO, 2005).

Historicamente, a noção de encontrar padrões úteis em dados inicialmente sem tratamento tem recebido diversos nomes, tais como: prospecção ou descoberta de conhecimento em bancos de dados, mineração de dados, extração de conhecimento, descoberta de informação, coleta de informação, arqueologia de dados ou processo de padronização de dados (BERNARDES NETO, 2001).

Dentre os conceitos apresentados, com base em Bretzke (2000) e Carvalho (2005), pode-se afirmar que o *Data Mining* não é uma tecnologia, mas um processo, com uma série de etapas inter-relacionadas, complementadas por *softwares* e técnicas, cujo objetivo final é descobrir informações ou padrões escondidos nos dados, que tenham valor significativo para as empresas.

Dependendo da abordagem, as definições de *Data Mining* têm maior ênfase na tecnologia, nas áreas de pesquisa da Estatística e da Inteligência Artificial ou no processo de descoberta de conhecimento (CORRÊA, 2007).

Neste trabalho, tendo como abordagem a gestão do conhecimento, o termo *Data Mining* será tratado como mineração de dados, mais comum na ciência da informação.

2.5.2 Histórico

O *Data Mining* surgiu devido a necessidade de transformar dados em informação.

Na década de 1980, ocorreu a popularização dos computadores e *softwares*. No início da era computacional, os computadores lidavam com dados puros e simples. Os programas de computador transformavam dados em informação, as quais os usuários podiam aplicar aos seus problemas de negócios específicos. Esses dados eram entendidos como uma determinada informação representada por uma seqüência de *bits* (PINHEIRO, 2003).

As inúmeras bases de dados de natureza comercial, administrativa, governamental e científica aumentaram devido a constante utilização de computadores, sistemas gerenciadores de banco de dados, leitores de códigos de barras, e dispositivos de memória secundária (GOLDSCHMIDT; PASSOS, 2005). As organizações passaram a coletar e reunir um grande conjunto de dados, num fluxo de informação sistemático sobre clientes e mercado (BRETZKE, 2000, p. 28).

A análise dos dados, por meio do conhecimento humano, já não era possível diante de grande volume de dados, e através dos sistemas existentes na década de 80. De acordo com Fayyad (1996 apud PAULA, 2004), as manipulações tradicionais de dados, através de *softwares* como planilhas e consultas *ad-hoc*, não seriam mais suficientes para o processamento de pesquisas em grandes bases de dados. Estes métodos podiam criar relatórios sobre os dados, mas não conseguiam analisá-los para destacar conhecimentos importantes. Assim, as organizações necessitavam de algum sistema inteligente e automático que complementasse o trabalho humano na tarefa de analisar, interpretar e relacionar esses dados para desenvolver estratégias de ação diante de um contexto observado (BERNARDES NETO, 2001).

A origem da mineração de dados foi propiciada pelos seguintes fatores: surgimento de potentes recursos computacionais; exigência de modernas técnicas de análise de dados para o apoio a decisão, e para conhecer a quem será oferecido os serviços; facilidades de aquisição de programas gratuitos (CARVALHO, 2005).

Com o surgimento de sistemas de recuperação da informação, os usuários descobriram que não queriam trabalhar com dados e sim com informações. Esta ferramenta

seria capaz de analisar, automaticamente, as bases de dados por meio de formulação de perguntas e, conseqüentemente, obter respostas rápidas (GOLDSCHMIDT; PASSOS, 2005).

Nesse novo cenário, os sistemas computacionais ganharam funcionalidades adicionais e, ao invés de apenas manterem as informações gerenciais para os usuários de negócio, passaram a disseminar informações gerenciais de forma atual, fornecendo grande poder de tomada de decisão a esses usuários (PINHEIRO, 2003).

Assim, os programas de computador passaram a transformar dados em informação, e os usuários passaram a aplicá-los em seus negócios específicos. As pessoas passaram a conhecer o seu conteúdo e interpretar os dados criando, assim, senso crítico e significado aos dados crus armazenados nos repositórios (PINHEIRO, 2003).

O termo Descoberta de Conhecimento em Base de Dados ou KDD (*Knowledge Discovery in Databases*) foi formalizado em 1989, em referência ao amplo conceito de procurar conhecimento em bases de dados (BERNARDES NETO, 2001).

2.5.3 Etapas de mineração de dados

De acordo com Goldschmidt e Passos (2005), a mineração de dado é uma técnica sistemática, não é realizada sem alguma metodologia. É caracterizada como um processo composta por 3 (três) etapas básicas: pré-processamento; mineração de dados, e pós-processamento.

2.5.3.1 Pré-processamento: organização e transformação dos dados

A etapa de pré-processamento compreende todas as funções relacionadas aos objetivos da aplicação, seleção, à organização e ao tratamento dos dados. Essa etapa tem como objetivo a seleção e a preparação dos dados para os algoritmos da etapa de Mineração de Dados (GOLDSCHMIDT; PASSOS, 2005).

Esta etapa visa preparar os dados para que possam ser processados pelas ferramentas de mineração de dados de modo a otimizar os resultados. É o processo que normalmente demanda a maior parte do tempo (CORRÊA, 2007).

De acordo com Carvalho (2005), o processo inicia-se com a definição clara do problema, ou seja, a definição do objetivo a ser alcançado. A definição de objetivos da aplicação compreende as características esperadas do modelo de conhecimento a ser

produzido ao final do processo. A definição de objetivos que a organização quer alcançar é realizada através de perguntas como: qual é a necessidade existente para executar a mineração de dados?; quais resultados a organização espera alcançar?; a organização tem recursos, tecnologia e usuários necessários para a utilização desta ferramenta?; será necessária uma experiência externa para complementar o grupo de trabalho?; que tipo de tarefas e técnicas de mineração de dados será mais adequada às análises que a organização precisa fazer? (TAURION, 2007).

Assim, em função do problema definido, escolhe-se a técnica (classificação, estimativa, previsão, etc.) e a ferramenta (redes neurais artificiais, algoritmos genéticos, etc.), capaz de executá-las (CARVALHO, 2005).

Os processos que envolvem esta etapa são:

- a) **Seleção:** a seleção dos dados compreende a identificação de quais informações, dentre as bases de dados existentes, devem ser efetivamente consideradas durante o processo de KDD. A seleção varia de acordo com interesses e resultados esperados e pode ter dois enfoques distintos: a escolha de atributos ou a escolha de registros. Diante da definição do problema a ser resolvido, escolhem-se os bancos de dados que serão alvos da pesquisa. Em muitos casos, as fontes de dados poderão estar em formatos diferentes, como planilhas eletrônicas, bancos de dados até mesmo em *Data Warehouse*. Para iniciar a etapa de mineração de dados, todas as fontes deverão estar reunidas em uma única base de dados ou planilha (GOLDSCHMIDT; PASSOS, 2005, p. 12). A maior diversidade das fontes de dados e dos formatos destes dados implicará em um maior esforço para reunir todas as fontes de dados. Decidir que atributos serão investigados no processo de mineração de dados é também parte da tarefa de seleção. Os atributos são os valores assumidos nos diversos registros dos arquivos (CORRÊA, 2007);
- b) **Limpeza dos dados:** a fase de limpeza dos dados envolve uma verificação da consistência das informações e consiste diminuir dados ruidosos e inconsistentes. Os dados ruidosos são dados errados ou que contenham valores divergentes, e dados inconsistentes são aqueles que contem alguma discrepância semântica entre si. A execução dessa fase tem como objetivo eliminar resultados inconsistentes que poderiam ser executados pelos algoritmos, afetando o desempenho destes (GOLDSCHMIDT; PASSOS, 2005, p.12). De acordo com Carvalho (2005), em uma

massa de dados podem ocorrer elementos ausentes por falha de digitação, erros de preenchimento de formulários, ou mesmo por que os registros pertencem a empresas ou épocas diferentes, nas quais um dado específico não era questionado ou considerado importante. Na complementação dos dados são definidos o valor padrão, definidos a priori. A limpeza dos dados (*data cleaning*), é realizada para que haja uma limitação da massa de dados a ser explorada, principalmente na redução do número de variáveis considerados na análise;

- c) **Codificação dos dados:** codificação de dados é a operação de pré-processamento responsável pela forma como os dados serão representados durante o processo de mineração, ou seja, como serão representados pelos algoritmos. Neste processo a codificação pode ser numérica- categórica, que transforma valores reais em categorias e intervalos; ou Categórica – Numérica, que representa numericamente valores de atributos categóricos (GOLDSCHMIDT; PASSOS, 2005, p.12) ;
- d) **Enriquecimento dos dados:** a tarefa de enriquecimento dos dados consiste em melhorar a informação contida nos registros dos bancos de dados através da criação de novos atributos a partir dos já existentes, agregando novas informações. A geração de totalizadores em variáveis numéricas, a criação de faixa ou classes de valores para atributos contínuos e a generalização de valores de atributos são exemplos de enriquecimento dos dados (CORRÊA, 2007). No processo de enriquecimento dos dados podem ser realizadas pesquisas para complementação dos dados, consultas a bases de dados externas, ou fontes de informação externa (GOLDSCHMIDT; PASSOS, 2005, p.12).

A etapa de preparação dos dados é realizada através de ferramentas de visualização e ferramentas de reformatação dos dados (*softwares de Data Mining*), o que corresponde a 60% do trabalho de DM. A preparação é determinante para a qualidade final dos resultados, por isso as ferramentas utilizadas são tão importantes (QUONIAM, 2001).

2.5.3.2 Mineração de dados: aplicação das técnicas e tarefas de mineração

De acordo com Goldschmidt e Passos (2005), a etapa de mineração de dados compreende a aplicação de algoritmos sobre os dados procurando abstrair conhecimento.

Durante a etapa de mineração de dados são definidos as técnicas e os algoritmos a serem utilizados no problema em questão. Estes algoritmos são fundamentados em técnicas que procuram, segundo determinados paradigmas, explorar os dados de forma a produzir modelos de conhecimento, ou seja, a forma de representação do conhecimento. A escolha da técnica depende do tipo de tarefa a ser realizado.

Na escolha de algoritmos, que serão utilizados para realizar alguma tarefa, vários modelos são estudados, mas apenas aqueles que se mostram mais hábeis para o encontro da solução desejada são desenvolvidos (GOLDSCHMIDT; PASSOS, 2005).

2.5.3.3 Pós-processamento: interpretação dos resultados e descoberta do conhecimento

A etapa de Pós-processamento abrange a visualização, análise, interpretação e a utilização dos resultados a partir do conhecimento obtido na Mineração de Dados. Nesta etapa, o conhecimento obtido através da mineração de dados é estruturado em forma de relatórios, gráficos, diagramas ou qualquer outra que apresente o conhecimento adquirido (GOLDSCHMIDT; PASSOS, 2005).

Cardoso (2005) observa que um relatório 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.

Em geral, nesta etapa o especialista de mineração de dados e o especialista em domínio de aplicação avaliam os resultados obtidos e definem novas alternativas de investigação dos dados (GOLDSCHMIDT; PASSOS, 2005).

As atividades de pós-processamento são:

- a) **interpretação dos resultados e descoberta do conhecimento:** na avaliação dos resultados ocorre uma avaliação criteriosa e a verificação de que as causas do problema foram sanadas ou o objetivo da empresa alcançado (CARVALHO, 2005). Nesta etapa é realizada a avaliação das saídas que foram fornecidas, mas apenas as informações úteis serão utilizadas pela organização. Os resultados

obtidos compreendem os modelos de conhecimento descobertos ao longo da aplicação de KDD e o histórico das ações realizadas (LIMA, 2006). De acordo com Fayyad *et al.* (1996a *apud* GOLDSCHMIDT; PASSOS, 2005), no modelo de conhecimento é avaliado o cumprimento das expectativas definidas nos objetivos de aplicação, realizada através de comparações de outros modelos de conhecimento obtido. O modelo de conhecimento com maior precisão na classificação dos clientes possui melhores chances de ser eleito como principal resultado;

- b) **uso das relações descobertas:** na fase de uso das relações descobertas, as decisões são tomadas de forma a utilizar da melhor forma possível as relações fornecidas pela mineração de dados (CARVALHO, 2005). Na análise dos resultados obtidos do processo de *Data Mining*, existem 2 (dois) aspectos fundamentais a serem considerados: informar as novas descobertas e apresentá-las de maneira que possam ser exploradas potencialmente. Nesta fase, recomenda-se a participação de um especialista da área de que trata a base de dados, para solucionar questões técnicas específicas que possam influir na análise. Gerentes de negócios e executivos podem ser envolvidos nesta fase (QUONIAM, 2001).

Eventualmente, após a avaliação dos resultados será necessário retornar aos outros ciclos para que as soluções propostas sejam eficientes. Assim, a primeira fase do processo deverá ser aplicada novamente, formando um ciclo de trabalho (CARVALHO, 2005).

As etapas de *Data Mining* assemelham-se com as etapas do ciclo informacional ou o processo de gestão da informação realizado dentro do domínio temático da ciência da informação, em especial no processo de recuperação de informações (*information retrieval*) (QUONIAM, 2001).

A figura abaixo mostra o modelo de mineração de dados proposta por Fayyad, Piatetsky - Shapiro e Smith , sendo muito comum na literatura sobre o assunto:

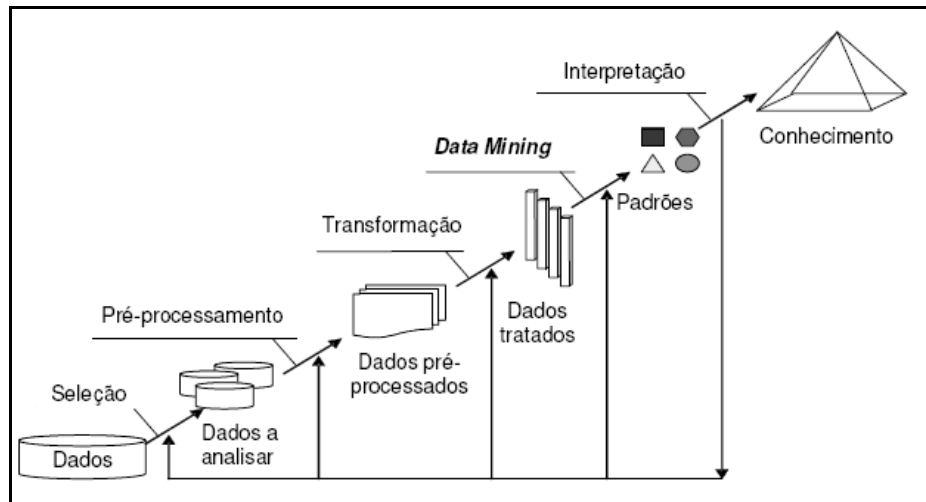


FIGURA 2– Etapas no processo de *Data Mining* de Fayyad, Piatetsky – Shapiro e Smith
 Fonte: Santos (2002 apud CARDOSO, 2005).

Um exemplo bem conhecido de utilização da mineração de dados é o caso da empresa de varejo dos Estados Unidos, a *Wal-Mart* na utilização de regras de associação. Durante a análise dos dados feita a partir das regras de associação, os resultados indicaram que quem compra fraldas para seus filhos também compravam cervejas para o consumo enquanto cuidavam das crianças e assistiam aos jogos na televisão durante o final de semana, com uma margem de erro. Após estas análises a *Wal-Mart* aumentou a venda de cervejas colocando-as próximas ao local onde as fraldas eram vendidas (GOLDSCHMIDT; PASSOS, 2005).

2.5.4 Tarefas de mineração de dados

A tarefa de mineração de dados, 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 tipos de padrões poderiam surpreender (AMO, 2003 apud CARDOSO, 2005).

2.5.4.1 Classificação

A classificação pressupõe que se conheçam as características que identificam os membros de uma determinada classe. Assim, a tarefa de classificação consiste em encontrar atributos e valores que melhor caracterizem cada classe, ou seja, identificar a qual classe que algum objeto pertence a partir de seu conteúdo (PAULA, 2004).

De acordo com Carvalho (2005), durante a classificação, alguns detalhes dos dados poderão ser desprezados e suas principais características examinadas sendo, posteriormente, atribuídas uma classe previamente definida. A tarefa de classificar normalmente exige a comparação de um objeto ou dado com outros dados ou objetos que, supostamente, pertençam a classes anteriormente definidas. A classificação utiliza um classificador binário: “verdadeiro” ou “falso”.

Esta técnica agrupa itens baseados em atributos pré-definidos como, por exemplo, identificar clientes em dia e inadimplentes com a empresa; identificar os clientes que moram na Barra da Tijuca, no Rio de Janeiro, que tenham apartamento próprio e viajam ao exterior (TAURION, 2007). Através dos dados históricos a empresa mapeará os clientes a partir de seus dados, em uma destas classes. Essa função pode ser incorporada a um sistema de apoio à decisão, que auxilie a filtragem e concessão de empréstimos somente a clientes classificados como bons pagadores (GOLDSCHMIDT; PASSOS, 2005).

As ferramentas utilizadas para a classificação são: Redes Neurais Artificiais; Estatísticas e Algoritmos Genéticos (CARVALHO, 2005). A classificação pode ser utilizada para classificar 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 (CARDOSO, 2005).

2.5.4.2 Estimativas

Enquanto a classificação lida com resultado discreto, a estimação trabalha com valor numérico contínuo (JESUS, 2004). Estimar algum índice é determinar seu valor mais provável diante dos dados do passado ou de dados de outros índices semelhantes sobre os quais se tem conhecimento (CARVALHO, 2005).

Dentre os exemplos de estimativa pode ser citada a análise de padrões de compra de determinado cliente para identificar se este tem filhos pequenos; estimar a renda de uma família (TAURION, 2007).

2.5.4.3 Previsão ou predição

A previsão é o mesmo que classificação ou estimação, exceto pelo fato dos registros serem classificados de acordo com alguma atividade futura prevista ou valor futuro estimado (JESUS, 2004).

A predição normalmente é utilizada quando se tem em mente um problema claro e bem especificado para ser resolvido, pois busca-se, através da mineração, uma resposta para este problema. O único meio de verificarmos se uma previsão foi bem feita é aguardar o acontecimento e verificar o quanto foi acertada ou não a previsão realizada. (SOUZA, 2003).

A previsão é uma das tarefas mais difíceis no *Data Mining*. A técnica da predição resume-se na avaliação do valor futuro de algum índice baseando-se em dados do comportamento passado deste índice. Inclui tarefas como determinar se o valor de uma ação de bolsa de valores subirá ou descera, ou qual será a população de uma certa cidade daqui há dez anos (CARVALHO, 2005).

2.5.4.4 Análise de agrupamento ou clusterização (segmentação)

A análise de agrupamento ou clusterização é utilizada para separar ou segmentar os registros de uma base de dados em subconjuntos ou *clusters*, onde os elementos compartilham propriedades comuns que os distingam de elementos em outros *clusters* (GOLDSCHMIDT; PASSOS, 2005).

Na análise de agrupamentos os grupos ou classes são construídos com base na semelhança entre os elementos, através da busca de correlações entre itens e indivíduos, deduzindo relacionamentos. Segmentar um mercado é uma típica análise de agrupamento, onde consumidores são reunidos em classes representantes dos segmentos deste mercado (CARVALHO, 2005).

As tarefas de agrupamento podem ser utilizadas em dados estruturados. Dados estruturados são dados bem definidos, pertencem a um domínio específico e apresentam relacionamentos e interações. As informações não estruturadas são definidas como aquelas sem uma organização previamente definida, e nesta categoria se enquadram as informações textuais (PAULA, 2004).

Como exemplo de agrupamento pode-se citar 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 (CARDOSO, 2005).

As áreas do conhecimento que tem utilizado as técnicas de agrupamento são: psicologia, arqueologia, medicina, genética e *marketing*, que se valem dos algoritmos de agrupamento para tarefas como organização de dados e recuperação de dados similares (PAULA, 2004).

2.5.4.5 Sumarização

Sumarização consiste em procurar identificar e indicar características comuns entre conjuntos de dados. Como exemplo de sumarização, um banco de dados com informações sobre clientes, que assinam um determinado tipo de revista semanal, poderá buscar por características que sejam comuns a boa parte dos clientes (GOLDSCHMIDT; PASSOS, 2005).

Como exemplo de sumarização tem-se uma base de dados de assinantes da revista X, que mostra homens na faixa etária de 25 a 45 anos, com nível superior e que trabalham na área de finanças. Estas informações poderiam ser utilizadas pela equipe de *marketing* da revista para direcionar a oferta para novos assinantes (GOLDSCHMIDT; PASSOS, 2005).

2.5.4.6 Descoberta de associação

A descoberta de associação tem como objetivo encontrar relacionamentos ou padrões freqüentes entre conjuntos de dados (GOLDSCHMIDT; PASSOS, 2005).

Esta tarefa objetiva determinar que “coisas” estão relacionadas, ou ocorrem juntos em um conjunto de dados, e se aplica nos casos em que se deseja estudar preferências, afinidades, visando criar oportunidades para formação de “pacotes” para consumidores (CORTES; PORCARO; LIFSCHITZ, 2002).

De acordo com Souza (2003), uma regra de associação é um padrão descritivo que representa uma declaração na forma X e Y. Como exemplo de descoberta de associação pode-se citar que 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 (CARDOSO, 2005).

2.5.5 Técnicas de mineração de dados

A técnica de mineração consiste na especificação de métodos que garantam “**como**” descobrir os padrões que interessam, ou seja, se referem a algum modelo que possa fundamentar a implementação de um método de KDD (AMO, 2003 apud CARDOSO, 2005).

Na aplicação de algoritmos são estabelecidos os padrões que, de acordo com Souza (2003), pode ser definido como sendo uma afirmação sobre uma distribuição probabilística. Estes padrões podem ser expressos principalmente na forma de regras, fórmulas e funções. Os padrões são encontrados através do auxílio nas buscas, utilizando diversas ferramentas, que podem ser aplicadas ao problema tanto isoladamente quanto em combinação. Através da aplicação destas ferramentas é gerado um conjunto de questões para que seja realizada a filtragem sobre as buscas em um dado e, a partir disto, enviar ao sistema os resultados obtidos (LIMA, 2006).

Após encontrar esses padrões, os algoritmos podem inferir regras, que podem então ser usadas para gerar um modelo que pode prever um comportamento desejado, ou então identificar um relacionamento entre dados, além de descobrir padrões ou fornecer critérios de classificação que possam auxiliar na formação de grupos de registros com similares de atributos. Por fim, o *Data Mining* gera um relatório com as descobertas feitas para que sejam estudados e interpretados (BERNARDES NETO, 2001).

2.5.5.1 Redes Neurais:

Redes Neurais ou RNA (Rede Neural Artificial) é uma técnica computacional que constrói um modelo matemático inspirado em um sistema neural biológico simplificado, com capacidade de aprendizado, generalização, associação e abstração (GOLDSCHMIDT; PASSOS, 2005, p. 17).

As redes neurais têm como pesquisador inicial o pesquisador Norbert Wiener que, na década de 1960, iniciou seus estudos sobre o processamento da informação nos seres vivos, criando a cibernética, originando a Engenharia Biomédica. Nos início dos anos 60, Rosenblatt, Minsky e Pappert mostraram que redes compostas dos neurônios poderiam aprender a classificar e reconhecer padrões visuais, fixando as bases do paradigma conhecido como Redes Neuronais Artificiais.

De acordo com Carvalho (2005), as redes neurais são compostas por duas estruturas: o nó, que corresponde ao neurônio; e o *link*, correspondente às conexões entre neurônios. As redes neurais são interessantes porque detectam padrões nos dados de forma análoga ao pensamento humano.

2.5.5.2 Algoritmos Genéticos

Algoritmos genéticos, como o próprio nome diz, são algoritmos que simulam o processo de seleção natural proposto por Charles Darwin, em 1859 (BERNARDES NETO, 2001).

Para Goldschmidt e Passos (2005), esta técnica utiliza de modelos de otimização, inspirados na evolução natural e na genética, aplicados a problemas complexos. As variáveis descrevem os vários aspectos do problema e as restrições indicam os valores que as variáveis podem ter.

Os algoritmos genéticos utilizam os operadores de seleção, cruzamento e mutação para desenvolverem sucessivas gerações de soluções (HARRISSON, 1998 apud JESUS, 2004). É considerado uma ferramenta poderosa para a mineração de dados. As deficiências dos algoritmos genéticos incluem a grande superprodução de soluções individuais, e elevada demanda de processamento computacional (CORTES; PORCARO; LIFSCHITZ, 2002);

2.5.5.3 Método estatístico: agrupamento e multivariada

De acordo com Gonçalves (2000), uma grande variedade de técnicas analíticas tem sido utilizada em mineração de dados. Existem técnicas mais tradicionais como a estatística multivariada, a estatística descritiva, a análise de agrupamentos e regressões, e modelos mais atuais de aprendizagem, dentre eles, as redes neurais, lógica difusa e algoritmos genéticos.

De acordo com Johnson (1998 apud FANDERUFF; MARTINS, 2002), a análise multivariada é uma metodologia composta por uma série de procedimentos e técnicas estatísticas definidas para extrair informações de situações que envolvem a avaliação simultânea de várias variáveis. O agrupamento ou cluster multivariado visa agrupar e discriminar grupos de indivíduos, regiões ou qualquer objeto em geral, que são constituídos utilizando-se baseados em distâncias. Segundo Alsabt (1999 apud FANDERUFF; MARTINS, 2002), a distância é uma medida matemática de semelhança, que pode ser geográfica,

temporal ou baseada em qualquer característica do objeto. Quando são utilizadas diversas variáveis é possível a construção de agrupamentos onde o critério de semelhança está distribuído entre diversas características.

A análise estatística de agrupamentos é uma ferramenta indicada para a descoberta não-supervisionada de conhecimento em mineração de dados, devido a simplicidade de seus algoritmos. Os algoritmos estatísticos para a análise de agrupamentos reproduzem os resultados dos mapas auto-organizáveis com a diferença de que estes últimos fornecem uma visão geométrica, às vezes muito útil (CARVALHO, 2005).

2.5.5.4 Árvores de Decisão

De acordo com Goldschmidt e Passos (2005), esse método é baseado na indução de árvore de decisão, aceito e utilizado pela comunidade científica mundial.

Carvalho (2005) define o método de árvores de decisão como um modelo de conhecimento, representado por uma árvore de busca, no qual uma sucessão de símbolos será gerada pela aplicação dos operadores. A partir da definição dos atributos é criada a árvore de decisão. Nesta técnica, cada nó interno denota um teste em atributo, cada ramo (sub-árvore) representa o resultado do teste e cada folha representa a distribuição dos registros. A partir dos dados de entrada, o processo da árvore de decisão executa uma seqüência teste juntamente com as possibilidades de cada nó (CORTES; PORCARO; LIFSCHITZ, 2002).

O algoritmo de árvore de decisão é formado por 2 (dois) tipos de atributos: o decisivo, que possui o resultado esperado; e os não-decisivos, que possuem valores que irão direcionar a uma decisão (LIMA, 2006).

Bernardes Neto (2001) afirma que “o principal problema com árvores é que elas eliminam dados em uma razão exponencial profunda. Desse modo, para cobrir as estruturas complexas, extensos conjuntos de dados são exigidos”.

2.5.5.5 Regras de associação

Análise de associação é o processo de interconexão de objetos na tentativa de expor características e tendências. Gera redes de interações e conexões usando as associações item a item. Na regra de associação, o banco de dados é visto como uma coleção de transações, cada uma envolvendo um conjunto de itens (CORTES; PORCARO; LIFSCHITZ, 2002).

Um exemplo comum é o referente à cesta do supermercado. Na área de marketing é conhecido como análises de transações de compras (*market basket analysis*) (CORTES; PORCARO; LIFSCHITZ, 2002).

2.5.6 Componentes de mineração de dados

O processo de *Data Mining* não é realizado apenas pela tecnologia, mas sim por um conjunto de elementos que juntos são responsáveis pelos resultados que serão obtidos.

De acordo com Goldschmidt e Passos (2005), o *Data Mining* é composta por: conjunto de dados, tecnologia, algoritmos e especialista do domínio da aplicação.

2.5.6.1 Conjunto de dados

O conjunto de dados refere-se aos dados que serão utilizados na mineração de dados. De acordo com Goldschmidt e Passos (2005), um conjunto de dados faz parte de uma estrutura ou esquema do conjunto de dados, contendo características ou atributos.

Existe uma série de fatores que devem ser observados nos dados, como tipo de dados e estrutura dos dados.

- a) **Tipo de dados:** os dados são fontes importantes para a mineração de dados e podem ser divididos em dados qualitativos e dados quantitativos. Os dados qualitativos são dados que contém palavras nominais ou categóricas, como por exemplo, perfis dos usuários ou dados demográficos. Os dados quantitativos são aqueles que expressam quantidades, medidas. Os dados qualitativos são geralmente utilizados em segmentação e classificação. Dependendo da finalidade de mineração de dados, o conjunto dos dados pode mudar e ser utilizados para a prospecção, avaliação de riscos e metas junto a clientes. Os dados qualitativos e quantitativos também podem ser transformados, ou seja, dependendo do *software* estes podem ser codificados de qualitativos para quantitativos e vice-versa. (BRAGA, 2004);
- b) **Estrutura dos dados:** na mineração de dados estes devem estar estruturados, ou seja, em estruturas fixas conhecidos como tabelas. Em geral, os dados precisam ser organizados em uma única estrutura tabular multidimensional, ou dispostas em tabelas para facilitar a análise categórica pelo *software* de mineração de dados. Neste

contexto, é importante observar como estão dispostas as variáveis, pois determinados algoritmos possuem restrições ou não são compatíveis. Por isso, pode haver a necessidade de eliminar certos atributos.

2.5.6.2 Tecnologia de mineração de dados

A tecnologia de mineração de dados envolve qualquer recurso computacional que possa ser utilizado para concepção, aprimoramento e desenvolvimento de algoritmos, processo de busca e análise de dados. Os recursos computacionais para a realização de mineração de dados englobam *hardware* (processadores de memória) e *softwares* (ferramentas de consultas) disponíveis para a execução da aplicação de KDD (GOLDSCHMIDT; PASSOS, 2005).

A primeira geração de ferramentas para mineração de dados apareceu nos anos 80, e consistia em ferramentas de análise voltadas a uma única tarefa, sem suporte às demais usando ferramentas de Indução de Regras ou de Redes Neurais, a descoberta de *clusters* (grupos) nos dados, ou ainda na visualização de dados. A segunda geração de sistemas de mineração de dados surge em 1995 com o desenvolvimento de ferramentas chamadas “*suítes*”, ou *softwares*, dentre eles o *SSPS*, *Clementine*, *Inteligente Miner* e *SSAS Enterprise Miner*, que permitiam ao usuário realizar tarefas de descoberta (geralmente classificação, clusterização e visualização) e suportavam a transformação de dados (PIATETSKY - SHAPIRO, 1999 *apud* GOLDSCHMIDT; PASSOS, 2005, p.7). A terceira geração de mineração de dados surge no final dos anos 90, quando essas soluções passam a ser orientadas para a resolução de um problema específico da empresa, como, por exemplo, detecção de fraudes em cartão de crédito. Nesta geração as interfaces são orientadas para o usuário e procuram esconder toda a complexidade da mineração de dados. A quarta geração de ferramentas de mineração de dados compreende o desenvolvimento e a aplicação de técnicas e ferramentas que auxiliem o homem na própria condução do complexo processo de KDD (GOLDSCHMIDT; PASSOS, 2005, p.7).

Atualmente existem no mercado diversos *softwares* tanto comerciais quanto livres, que implementam ambientes integrados para facilitar a execução das etapas operacionais. Em geral, estas ferramentas reúnem diversos métodos de mineração de dados podendo ser aplicadas em várias tarefas (GOLDSCHMIDT; PASSOS, 2005).

As principais ferramentas de mineração de dados são:

- a) **Weka (*Waikato Environment for Knowledge Analysis*)**: a ferramenta Weka, foi criada pela Universidade de Waikato na Nova Zelândia. É uma ferramenta que trabalha diversas técnicas de *Data Mining*, e foi desenvolvida utilizando a linguagem Java. O *software* é de fácil instalação e manuseio, e tem sido uma das mais utilizadas no processo de mineração de dados (WITTEN; FRANK, 2000 *apud* LIMA, 2006). O Weka suporta a abertura direta de arquivos ARFF, CSV (Separados por vírgulas),. O formato ARFF é um arquivo ASCII usado para definir atributos a seus valores. A visualização dos dados no Weka se dá em forma de histogramas. A apresentação dos resultados pode ser obtidos em forma de árvores de decisão, e modelos gráficos para montagem de redes neurais (GOLDSCHMIDT; PASSOS, 2005);
- b) **SPSS Clementine (*Statistical Package for the Social Sciences*)**: é um *software* que permite o gerenciamento e a análise estatística dos dados. O SPSS é formado por uma família de produtos, que podem ser adquiridos e utilizados de forma independente, com destaque especial para o Clementine, muito utilizado para a análise de dados comerciais e científicos. O SPSS possui uma estrutura de organização de dados própria, mais permite a importação de dados em diversos formatos, tais como: FoxPro, Access, Dbase, ACSII, XLS (Excel), *Oracle Server* (GOLDSCHMIDT; PASSOS, 2005);
- c) **SPAD (*Système Pour Analyse de Données* ou Sistema Portátil de Análise de dados)**: o *software* SPAD é uma ferramenta de análise de dados e de *data mining*. O SPAD contém todos os métodos de análise estatística e análise exploratória de dados simples e multidimensional, permitindo uma análise múltipla dos dados de forma simultânea (CAMPOS, 2003).

Segundo Morzy et al. (2000 *apud* CARDOSO, 2005), as ferramentas de análise podem ser baseadas em algoritmos ou em visualização (computação gráfica). 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.

2.5.6.3 Algoritmos

O algoritmos são uma seqüência de operações matemáticas, precisamente definidas, capazes de realizar operações complexas, gerando programas que executam tarefas “inteligentes”. As diversas técnicas de mineração de dados utilizam vários algoritmos na solução dos problemas, muitos desses algoritmos estão disponíveis em *softwares* matemáticos e estatísticos através de pacotes que possuem tarefas específicas de acordo com o problema que se deseja solucionar (CUNICO, 2005).

Os algoritmos de mineração mais comuns são: Algoritmos Estatísticos, Algoritmos Genéticos, Árvores de Decisão, Redes Neurais Artificiais, e Algoritmos de Agrupamento (CARVALHO, 2005).

2.5.6.4 Especialista do domínio da aplicação

Representa a pessoa ou o grupo de pessoas que conhece o assunto e o ambiente em que deverá ser realizada a aplicação de *Data Mining*. O especialista em mineração de dados costuma deter o chamado conhecimento prévio sobre o problema, e influenciam na definição dos objetivos do processo até a avaliação dos resultados. O analista humano tem a tarefa de orientar a execução do processo de *Data Mining*, utilizando de suas experiências, seus conhecimentos e sua intuição para interpretar e combinar subjetivamente os fatos, de forma a decidir qual a estratégia a ser adotada. O conhecimento deste técnico é insubstituível, para executar as tarefas e técnicas de mineração de dados. (GOLDSCHMIDT; PASSOS, 2005).

A mineração de dados é multidisciplinar e origina de diversas áreas como: estatística, inteligência computacional, aprendizado de máquina, reconhecimento de padrões, banco de dados, podendo incluir a gestão do conhecimento (GOLDSCHMIDT; PASSOS, 2005).

A mineração de dados é uma área de pesquisa que necessita de uma equipe multidisciplinar, incluindo principalmente áreas de 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 dado (CARDOSO, 2005).

2.5.7 Áreas de aplicação de mineração de dados

O *Data Mining* tem-se tornado uma ferramenta importante para o apoio a decisão nas organizações privadas e públicas, sendo utilizado para a mineração de dados “escondidos” de grande base de dados (CORTES; PORCARO; LIFSCHITZ, 2002).

A cada dia os dados são gerados quando um cliente realiza uma compra por telefone, por exemplo; com o registro de duração da chamada de um número telefônico; quando alguém utiliza o cartão de crédito; quando um cliente preenche seus dados como nível sociocultural, preferências e *hobbies*, podem ser facilmente adquiridos e armazenados em bancos de dados (CARVALHO, 2005).

A utilização do *Data Mining* pode ser diversa, e de acordo com a necessidade e o objetivo da organização:

- a) **Varejo:** grandes grupos na área de varejo utilizam *Data Mining* para estudar o comportamento de compra de seus clientes. Através do cadastramento de clientes com um cartão específico que, utilizado no momento da compra, identifica as características pessoais do cliente, tais como: sexo, idade, estado civil, etc; e as características dos produtos adquiridos. Como resultados um supermercado poderá organizar suas gôndolas de acordo com os padrões de compras associadas entre produtos (CUNICO, 2005);
- b) **Sistema financeiro:** analisar empréstimos baseados nos padrões de compra, crédito e endividamento dos clientes com o objetivo de prever quais se tornarão inadimplentes (TAURION, 2007);
- c) **Telecomunicações:** uma empresa de telefonia celular poderá, baseada em padrões de uso dos telefones e outros fatores de mercado, fazer previsões de quais clientes estão propensos a migrar para as concorrentes (TAURION, 2007);
- d) **Pesquisas científicas:** as técnicas de mineração de dados podem ajudar cientistas em suas pesquisas, por exemplo, encontrar padrões em estruturas moleculares, dados genéticos, mudanças globais de clima, etc. (JESUS, 2004);
- e) **Coordenação de projetos científicos:** para os projetos de pesquisas em universidades, os resultados da mineração de dados podem propiciar a coordenação de

pesquisa obter informações que serão úteis para a administração dos projetos e até mesmo quanto aos investimentos feitos, auxiliando nas tomadas de decisões a serem realizadas. Pode ser um meio de analisar o desempenho destes trabalhos e o interesse de cada curso no desenvolvimento destes projetos (LIMA, 2006).

- f) **Unidades de informação:** denominado de *bibliomining* na Europa e Estados Unidos, foi cunhado por Nicholson e Stanton da Universidade de Syracuse, nos Estados Unidos. Em unidades de informação, a mineração de dados pode ser utilizada para conhecer padrões de comportamento de seus usuários, para verificar a utilização de determinado documentos, para analisar a eficiência ou deficiência nos sistemas de busca *online*, permitindo o melhoramento nos produtos e serviços oferecidos pela unidade de informação. (NICHOLSON, 2004).

A análise de gigantescos bancos de dados através de algoritmos computacionais vem auxiliando atualmente governos, exércitos, universidades, empresas e pessoas na obtenção de valiosas informações que auxiliam significativamente na tomada de decisão, minimizando erros, custos e tempo e maximizando acertos, lucros, rapidez e satisfação (CUNICO, 2005).

A mineração de dados busca identificar perfis dos diferentes usuários ou clientes e seus diversos padrões de consumo ou comportamento. Trabalhando por segmentos, podem-se identificar o perfil dos melhores clientes; os produtos e os serviços consumidos, as características de cada segmento de clientes; como e quando dois ou mais produtos ou características se associam ou se agrupam numa única compra ou evento e também o padrão de consumo/ comportamento de uma região, bairro, idade ou sexo (PENTEADO, 2006).

3 METODOLOGIA

Esta seção tem como objetivo descrever a metodologia e o método utilizado para o processo de mineração de dados, a fim de analisar o perfil de grupo de pesquisas da UFG.

De acordo com Ruiz (2002), a palavra “método” é de origem grega e significa o conjunto de etapas e processos a serem vencidos, ordenadamente, na investigação dos fatos ou na procura da verdade. Pode, ainda, significar uma série de etapas para se tentar resolver um problema. Entende-se por método, no sentido geral, o conjunto de atividades sistemáticas que, com maior segurança e economia, permitem atingir determinado objetivo.

3.1 NATUREZA DA PESQUISA

Este estudo tem a natureza exploratória e descritiva.

De acordo com Sekaran (1992 apud FERNANDES, 2007), o estudo descritivo tem a pretensão de descrever e compreender as características de variáveis que um estudo possui. A pesquisa descritiva descreve situações e eventos, visando dizer como é e como se manifesta um determinado fenômeno. A pesquisa descritiva é marcada por um enunciado claro do problema, possui hipóteses específicas e a necessidade de informações bem detalhadas.

Malhotra (2006) define pesquisa exploratória como a utilizada em casos nos quais é preciso definir o problema com maior precisão, identificar cursos relevantes de ação ou obter dados adicionais antes de poder desenvolver uma abordagem. Na pesquisa exploratória o pesquisador não possui um entendimento suficiente para prosseguir com o projeto de pesquisa, é mais flexível e versátil permitindo alterações durante a pesquisa até que as possibilidades sejam esgotadas ou outra direção seja descoberta.

3.2 MÉTODOS UTILIZADOS

3.2.1 Procedimentos de pesquisas

Os procedimentos utilizados para a realização deste estudo foram a pesquisa bibliográfica e o estudo de caso.

A pesquisa bibliográfica envolveu a leitura, análise e consultas as fontes primárias como livros e artigos eletrônicos, com o intuito de verificar as diversas definições, objetivos, características e abordagens teóricas de diversos autores. Os temas da pesquisa bibliográfica foram referentes a pesquisa científica, gestão do conhecimento, sistemas de gestão do conhecimento e mineração de dados.

No estudo de caso foi levantado, através da pesquisa descritiva, os pontos fortes e pontos fracos a fim de entender e descrever esses pontos dos resultados gerados da análise dos dados por meio da mineração de dados. Neste estudo adotou-se técnicas utilizadas na mineração de dados envolvendo processos como: identificação dos objetivos; obtenção e entendimento dos dados; pré-processamento dos dados; mineração de dados; e análise dos resultados.

As técnicas e as tarefas utilizadas para a mineração de dados estão descritas na análise dos dados, mas pode ser visualizadas na seguinte figura:

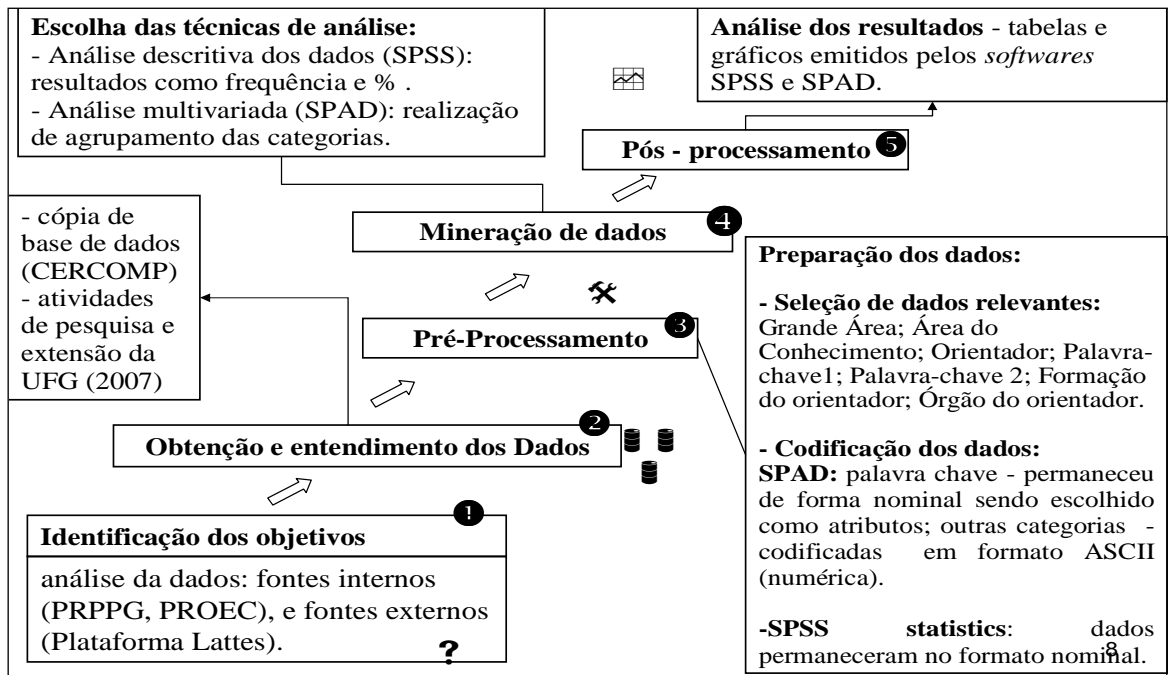


FIGURA 3 - metodologia de mineração de dados utilizados para o estudo de caso
Fonte: elaborada pela autora (2008)

A análise dos dados foi realizada em uma cópia de dados contendo dados que são cadastrados pelos órgãos responsáveis pela coordenação de pesquisa da universidade como a PROEC (Pró-Reitoria de Extensão e Cultura), PRPPG (Pró-Reitoria de Pesquisa e Pós-Graduação) e SICAD (Sistema de Cadastro das Atividades Docentes).

3.2.2 Instrumentos de pesquisa

O instrumento de pesquisa consistiu numa análise de uma cópia sobre as atividades de pesquisa da UFG. Consistiu na etapa de estudo do ambiente tecnológico dos dados.

As principais categorias utilizadas foram: grande área; área do conhecimento; sub-área do conhecimento; especialidade; orientador; palavra-chave; formação do orientador; órgão do orientador.

As ferramentas utilizadas para a manipulação dos dados foram:

- **Software SPSS statistics 17.0:** utilizado para a análise descritiva dos dados;
- **SPAD (Sistema Portátil de Análise de Dados):** versão 5.0, utilizado para a análise multivariada dos dados;

- **Base de dados:** sobre cadastro de pesquisa da UFG;
- **Planilha Microsoft Excel:** para a organização dos dados;

Os testes foram realizados em um micro-computador com 256 MB de memória RAM e sistema operacional Windows XP.

3.3 CONTEXTO DA PESQUISA

3.3.1 Descrição da instituição

Criada em 1960, a UFG tem investido nos seus 48 (quarenta e oito) anos na sua expansão, tanto na estrutura física quanto na estrutura de ensino.

A UFG possui, ao todo, mais de 31.000 m² de área construída, que correspondem a cerca de 17% do espaço edificado de toda a UFG, estimado em 213.000 m² de área. A UFG possui 27 (vinte e sete) unidades incluindo: o Centro de Ensino e Pesquisa Aplicado à Educação (CEPAE), com a função de ministrar o ensino fundamental e médio; e seis órgãos suplementares: Biblioteca Central, Centro Editorial e Gráfico, Centro de Processamento de Dados, Hospital das Clínicas, Museu Antropológico e Rádio Universitária. Além disso, possui 5 (cinco) campus no interior do Estado.

A UFG é formada por um vasto capital humano, que são representados pelas competências acadêmicas, técnicas e gerenciais. O capital humano da UFG é formado por docentes (professores com dedicação exclusiva e professores substitutos), e profissionais do quadro técnico administrativo (funcionários de nível superior, médio e de primeiro-grau).

Criado em 10 de setembro de 1934, e instalado em 20 de agosto de 1997, a Faculdade de Comunicação e Biblioteconomia (FACOMB), têm realizado investimentos em pesquisas e produção do conhecimento nas áreas de Comunicação e Biblioteconomia. O corpo docente da FACOMB conta com cerca de 29 (vinte e nove) professores sendo 11 (onze) com nível de doutorado, 14 (quatorze) com título de mestrado e 4 (quatro) sem titulação.

3.3.2 Produção científica da UFG

A UFG tem contribuído para a geração, sistematização e socialização do conhecimento e o saber, e formando profissionais e indivíduos capazes de promover a transformação e o desenvolvimento da sociedade.

A socialização do conhecimento e o saber, é oferecida pela instituição, através das atividades de ensino e pesquisa oferecidos pelos cursos de Graduação, Especialização e Pós-Graduação (Mestrado e Doutorado), em diversas áreas do conhecimento. A UFG conta com 695 (seiscentos e noventa e cinco) linhas de pesquisa, distribuídas em 229 (duzentos e vinte e nove) grupos de pesquisa em atividade.

3.4 ANÁLISE DOS DADOS

Para o estudo de caso foram utilizadas as seguintes etapas de mineração de dados: identificação dos objetivos; obtenção e entendimento dos dados; pré-processamento dos dados; mineração de dados; e análise dos resultados.

3.4.1 Identificação dos objetivos

A definição dos objetivos teve como pressuposto que nos registros da pesquisa científica da UFG, existe uma grande quantidade de informação e conhecimento “escondidos”. A mineração dos dados permitiria identificar o perfil e competência dos orientadores e das atividades de pesquisas da UFG que é o foco do estudo.

Para a identificação do objetivo da mineração de dados foram realizadas pesquisas nas páginas da PROEC, PRPPG³, e nas cópias de base de dados destes sistemas, para entender a estrutura de cadastro da produção científica realizada na UFG. A partir deste estudo foi possível verificar como as atividades de pesquisas da UFG são distribuídas e organizadas.

³ As pesquisas foram realizadas durante o mês de Agosto e Setembro do ano de 2008, quando estavam disponíveis nas seguintes páginas: www.prppg.ufg.br; www.proec.ufg.br. Atualmente não estão disponíveis devido a mudanças ocorridas nas páginas pesquisadas.

3.4.1.1 Análise do Sistema de Cadastro da UFG

A UFG possui, como estrutura de cadastro, os sistemas de cadastro da PRPPG (Pró-Reitoria de Pesquisa e Pós-Graduação), a PROEC (Pró-Reitoria de Extensão e Cultura), e o SICAD (Sistema de cadastramento de atividades docentes).

- a) **Sistema de cadastro de atividades docentes (SICAD):** é o sistema de cadastramento de atividades docentes. Neste sistema, todo professor lança todas as suas atividades de ensino (disciplinas ministradas na graduação e pós-graduação), pesquisa (todos os seus projetos de pesquisa), extensão, administrativas (participação em conselhos, coordenações, etc.) e publicações. O SICAD é mais completo e atualizado que o PRPPG e o PROEC, sendo atualizado uma vez por ano, em meados de dezembro;
- b) **Sistema de cadastro da pró-reitoria de pesquisa e pós-graduação (PRPPG):** a Pró-Reitoria de Pesquisa e Pós-Graduação (PRPPG) é subordinada diretamente a Reitoria, é responsável pela coordenação, expansão e incentivo a pós-graduação da UFG, principalmente nos cursos de mestrado e doutorado. O sistema utilizado pela PRPPG é o Sistema de Acompanhamento de Projetos de pesquisa (SAPP). Com este sistema, o setor de cadastro de pesquisa da PRPPG destina-se a colher informações sobre projetos de pesquisa desenvolvidos pelos docentes desta IES. Os campos utilizados para os dados de pesquisa cadastrado na PRPPG são: docente, matrícula (UFG), número do projeto, título, início, término, nome do participante, local.
- c) **Sistema de cadastro da pró-reitoria de extensão e cultura (PROEC):** a UFG por intermédio da PROEC, cadastra, coordena e oferece suporte técnico e de divulgação para as políticas de extensão e cultura da comunidade universitária. A PROEC⁴ mantém cadastro de extensão e cultura (cursos, consultorias, eventos, etc.). No ano de 2007 a UFG dispunha de 431 (quatrocentos e trinta e um) projetos de extensão e cultura universitária. A extensão universitária é um serviço prestado pela universidade a população, através da participação dos discentes. Uma das atividades da PROEC é divulgar o catálogo de extensão com as ações desenvolvidas pela UFG. Os dados da PROEC contêm as seguintes estruturas: código da faculdade, título original, título

⁴ Até o momento da pesquisa, realizada em Agosto e Setembro de 2008, as informações sobre atividades de extensão e cultura da UFG podiam ser obtidas através da página eletrônica da PROEC (<http://www.proec.ufg.br/>), fornecidas pelo SIEC. As informações eram obtidas ao pesquisar no campo 'Buscar ação de extensão'.

atual, unidade / órgão responsável, grande área de conhecimento, área temática primária, área temática secundária, palavras-chaves, atividade, público alvo, local de realização, início das atividades, término previsto, situação, coordenador, ementa, e-mail do coordenador.

Através da análise dos sistemas de cadastro da UFG, os objetivos identificados para a realização de mineração de dados foram:

- Quantas pesquisas são realizadas por Grandes áreas do conhecimento?;
- Quantas pesquisas são realizadas por Áreas do conhecimento?;
- Qual é o nível de formação dos orientadores? ;
- Quais são os órgãos/Unidades a que os orientadores pertencem?;
- Quais são os tipo de participantes que mais ocorrem nas atividades de pesquisa?;
- Como estão distribuídos os participantes em relação às Grandes Áreas do conhecimento?
- Quais são as palavras-chaves mais utilizadas pelos orientadores?.

3.4.2 Obtenção e entendimento dos dados

De acordo com Shearer (2000 apud FERNANDES, 2007), a compreensão dos dados é importante para que se obtenha um maior conhecimento e familiaridade a respeito de sua natureza, da qualidade presente nestes dados, bem como a necessidade de dados adicionais.

Os dados sobre a produção científica da UFG foram obtidos por meio de um pedido de cópia dos dados sobre atividades cadastradas pelo sistema de cadastro da UFG, solicitados a CERCOMP⁵ (Centro de Recursos Computacionais).

⁵ Centro de Recursos Computacionais da UFG, órgão responsável pela administração da tecnologia de informação e sistemas de banco de dados da universidade.

3.4.3 Pré-processamento: metodologia de mineração de dados

A etapa de pré-processamento é uma das etapas que mais demanda tempo e paciência por parte dos especialistas, pois compreende todas as funções realizadas a preparação dos dados como a seleção, limpeza, construção, a organização, integração e ao tratamento dos dados.

Neste estudo, os processos utilizados para o pré-processamento dos dados foram a seleção e transformação dos dados.

Após a coleta inicial foi realizada a etapa de exploração dos dados para descrevê-los e analisar a quantidade de registros e de campos presentes nas bases de dados. Estes dados estavam dispostos em 3 (três) tabelas no formato Excel, mas foi escolhida 1 (uma) tabela devido ao melhor aproveitamento. A tabela escolhida estava estruturada e distribuída em 16 (dezesesseis) atributos, continha 5784 (cinco mil, setecentos e oitenta e quatro) registros. Para a análise dos dados no SPSS *statistics* 17.0 e SPAD foram escolhidas as seguintes categorias: Grande Área, Área do conhecimento, Orientador, Órgão do orientador, Formação do orientador, Especialidade, Sub-Área, Nome do participante, Palavra chave 1, Palavra chave 2.

No *software* SPSS *statistics*, a planilha foi importada com o formato original em Excel para a interface que permite a organização e análise dos dados. Neste *software* os dados não foram codificados, permanecendo no formato nominal (sem codificação numérica).

No *software* SPAD, a planilha foi salva em formato CSV (Separado por vírgulas), pois o SPAD depende deste formato para a importação dos dados. Posteriormente os dados contidos na planilha foram codificados em formato ASCII (numérica), para facilitar a leitura do *software* SPAD.

Os processos utilizados para a codificação dos dados no SPAD foram: seleção das palavras para a análise, ou seja, as categorias; construção dos seguimentos de palavras-chaves; e a elaboração do agrupamento das palavras-chaves nos vetores, fazendo com que os dados sejam mostrados de forma condensada ou em forma de cluster/agrupadas.

O processo de codificação dos dados no SPAD, pode ser verificado no Anexo B, na figura contendo a tela de edição utilizada para a codificação dos dados.

3.4.4 Mineração de dados

Para o processo de mineração de dados foi escolhido o método estatístico, sendo utilizados 2 (dois) tipos de atividades para a análise de dados: a análise descritiva e a análise multivariada.

A análise descritiva se deu através do *software SPSS statistics 17.0*. Este *software* foi importante devido a sua capacidade de analisar os dados contidos em tabelas no formato Excel de forma automática e simultânea, facilitando o trabalho de construção de tabelas. A análise no SPSS ocorreu de duas formas: por meio da análise de frequência e através da correlação entre as categorias. Na análise de frequência foram analisados as categorias de forma individual, permitindo a obtenção da quantidade em porcentagem que ocorrem em cada valores /categoria. A análise por meio da correlação entre as categorias gerou regras de associação, obtendo assim resultados de forma agrupada.

A análise multivariada foi realizada análise por agrupamento ou clusterização com o *software SPAD (Sistema Portátil de Análise de Dados)* na versão 5.5. A análise por agrupamento buscou identificar assuntos similares que ocorrem nas variáveis escolhidas.

O SPAD tem como características colocar os valores de forma simultânea em forma de matriz, em até 11 (onze) eixos. Neste estudo foi utilizado 3 (três) eixos para facilitar a visualização e o agrupamento das classes. As grandes áreas foram agrupadas em três classes para uma melhor análise do *software*, e a distribuição nos eixos da matriz. Os detalhes desta representação pode ser verificada no quadro abaixo, e na figura 04.

NÚMERO	GRANDE ÁREA
01	Ciências Agrárias
02	Ciências Biológicas
03	Ciências da Saúde
04	Ciências Exatas e da Terra
05	Ciências Humanas
06	Ciências Sociais Aplicadas
07	Engenharias
08	Outros

QUADRO 01 – Número das Grandes Áreas na matriz

CLASSES	GRANDES ÁREAS
1/3	Ciências Agrárias Ciências da Saúde Ciências Humanas Linguística, Letras e Artes Engenharias Ciências Sociais Aplicadas
2/3	Ciências Exatas e da Terra
3/3	Ciências Biológicas

QUADRO 02 – Representação das Classes e Grandes Áreas

3.4.5 Pós – processamento: interpretação dos resultados

Os resultados da mineração de dados foram respectivamente, obtidos a partir de relatórios como tabelas e gráficos emitidos pelos *softwares* SPSS e SPAD.

Na análise por meio da estatística descritiva obteve-se resultados como a frequência dos dados, e a porcentagem. Este *software* gerou tabelas e gráficos contendo as características destas, complementando o processo de mineração de dados.

Os resultados fornecidos pelo SPAD não foram em porcentagens, mas em forma de *tags*, ou seja, as variáveis foram distribuídas em nuvens de palavras, que podem ser visualizadas nos eixos da matriz.

Estes relatórios foram analisados e interpretados, e os resultados estão na seção resultados da pesquisa.

4 RESULTADOS DA PESQUISA

Esta seção tem como principal objetivo mostrar os resultados obtidos das análises dos dados.

O pressuposto inicial da pesquisa era a possibilidade de existir uma grande quantidade de informação e conhecimento “escondidos” nos registros da pesquisa científica da UFG. Através das técnicas e tarefas de mineração este conhecimento seria obtido.

4.1 ANÁLISE DESCRITIVA DOS DADOS

A análise descritiva dos dados foi realizada através do *software SPSS statistics 17.0*. Neste *software* por meio de análise de frequência obtive resultados mostrando a frequência de cada questão dos objetivos analisados, mas com ênfase a porcentagem. Os resultados da frequência estão no anexo A deste trabalho.

Os objetivos da análise foram encontrar: Quantidade de pesquisas por Grandes áreas do conhecimento; Quantidade de pesquisas por Áreas do conhecimento; Nível de formação dos orientadores e Tipo de participante.

4.1.1 Quantidade de pesquisas por grandes áreas do conhecimento

Em relação às grandes áreas do conhecimento o objetivo foi saber a quantidade de pesquisas existentes.

As grandes áreas do conhecimento da UFG analisados foram: Ciências Agrárias, Ciências da Saúde, Ciências Humanas, Lingüística, Letras e Artes, Ciências Sociais Aplicadas, Engenharias, Ciência da Computação, Ciências Exatas e da Terra e Ciências Biológicas.

No gráfico 1, os resultados mostram que o número de atividades de pesquisa está concentrado na maior parte nas Ciências Agrárias, com 25, 6%. Em relação às outras áreas, o número de pesquisas existentes no ano de 2007 são: Ciências da Saúde com 23,6%; Ciências Humanas 15,1%; Ciências Biológicas com 11,8%; Ciências Exatas e da Terra com 9,5%; Lingüística, Letras e Artes com 6,6%; Ciências Sociais Aplicadas com 4%, Engenharias com 3,3%. Assim, as áreas que tiveram maior participação no projeto de pesquisa científica e extensão foram as Ciências Agrárias e as Ciências da Saúde.

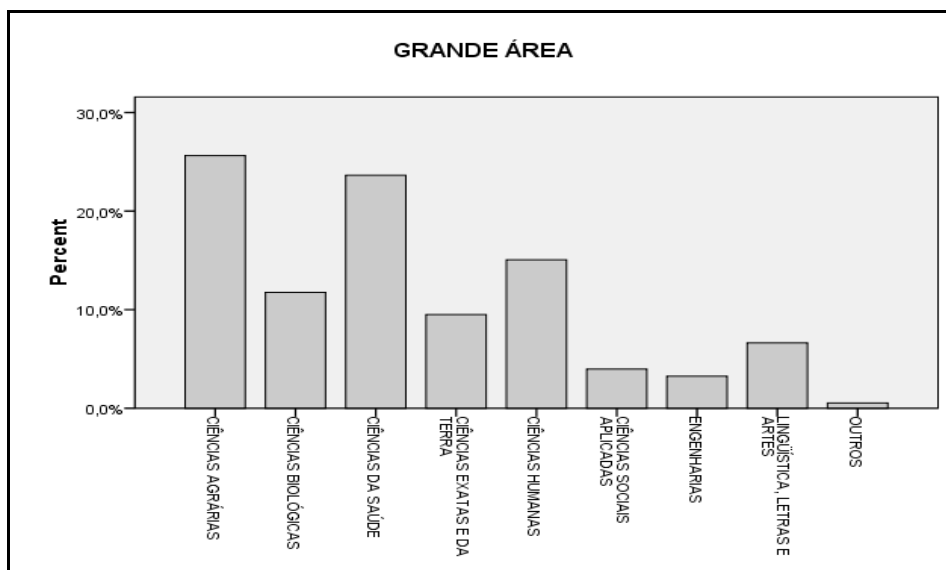


GRÁFICO 1 - Quantidade de pesquisas por Grandes áreas do conhecimento
 Fonte: elaborado pela autora

4.1.2 Quantidade de pesquisas por Áreas do conhecimento

No gráfico 2 mostra os dados sobre a quantidade de pesquisa por áreas do conhecimento.

Os dados mostram que Agronomia tem a maior porcentagem quanto às atividades de pesquisas com 10,2%; seguida de Medicina com 6,5%; Medicina Veterinária com 6,3%; Zootecnia com 5,1%; Educação com 5,2%; Zootecnia com 4,7% e Odontologia com 3,7%. O fato de cursos tradicionais na UFG possuírem a maior quantidade de pesquisas se refere ao fato de possuir mais quantidade de doutores e mestres, como verificados no gráfico 6, e por serem dependentes de pesquisas. A área de Ciência da Informação conta apenas com 0,2% das atividades de pesquisa.

As áreas como Agronomia, Medicina e Medicina Veterinária, recebem atenção especial por parte do governo federal, pois, muitos resultados de pesquisas beneficiam em grande parte a população brasileira e as organizações que utilizam os resultados proporcionados pela pesquisa científica.

Nessas áreas grande parte das pesquisas são sobre assuntos relacionados ao tratamento e saúde coletiva, produção animal, agronegócios e tecnologia de alimentos.

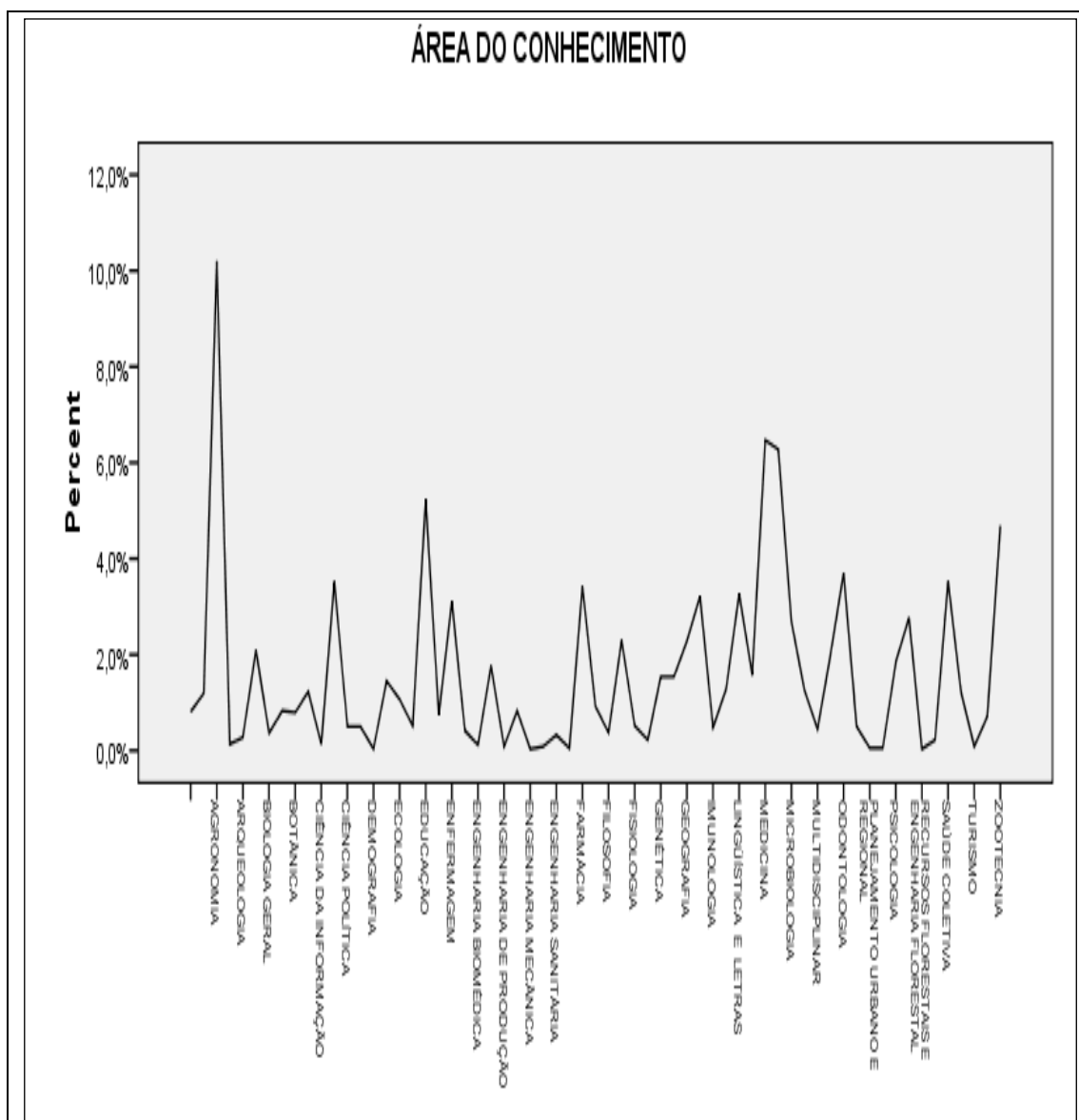


GRÁFICO 2 – Quantidade de pesquisas por Áreas do conhecimento

Fonte : elaborado pela autora

4.1.3 Nível de formação dos orientadores

Os resultados no gráfico 3 mostram que 84,9% das pesquisas existentes são coordenadas por doutores; 13,2% das pesquisas são coordenadas por mestres; 1,7 % das pesquisas são coordenadas por orientadores com formação em Graduação; e 0,2 % das pesquisas são coordenadas por orientadores com titulação em especialização.

Este resultado reflete os dados do ano de 2007, divulgados pela universidade do qual aponta que dos 1657 professores, 741 possuem o título de doutorado e 429 possuem mestrado.

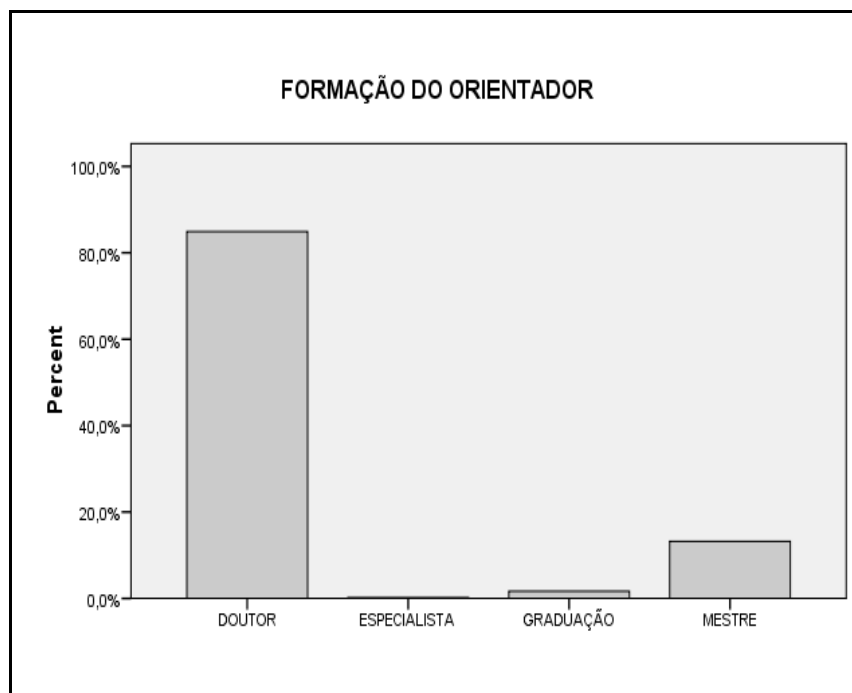


GRÁFICO 3 – Nível de formação dos orientadores
 Fonte: elaborado pela autora

4.1.4 Número de pesquisas existentes por órgãos e unidade acadêmica

O estudo tem como objetivo saber quais são os órgãos ou unidades acadêmicas que os orientadores pertencem.

O gráfico 4 mostra que 14,8% das pesquisas são realizados por orientadores do Campus de Jataí; 10,5% das pesquisas pertencem aos orientadores da escola de agronomia e engenharia de alimentos; 10,4% dos orientadores pertencem a escola de veterinária; 7,4% dos orientadores pertencem ao Campus de Catalão; 6,7% dos orientadores pertencem ao instituto de ciências biológicas. A Faculdade de Comunicação e Biblioteconomia possui cerca de 0,6 % dos orientadores presentes nas atividades de pesquisas realizadas na UFG.

Comparando os gráficos 3 e o gráfico 4, percebe-se que as unidades com maior índice de doutorado e mestrado possuem o maior índice de pesquisa, e os centros com menor percentual de docentes apresentaram menor percentual de pesquisa.

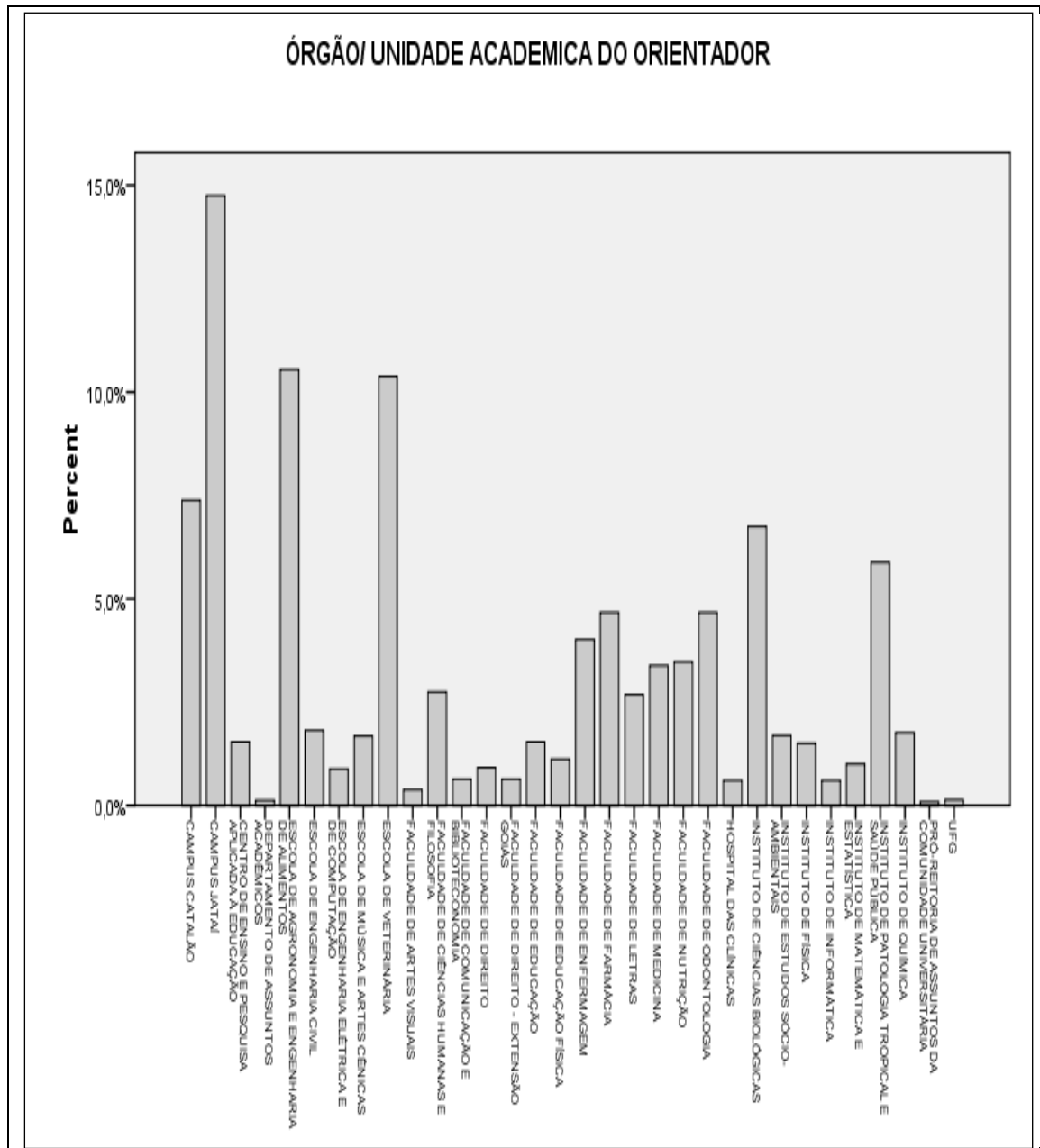


GRÁFICO 4 – Pesquisa realizadas por órgãos/unidades acadêmicas dos orientadores
Fonte: Elaborado pela autora

4.1.5 Tipo de participante

Quanto ao tipo de participante, o objetivo foi saber qual é o perfil dos pesquisadores. O gráfico 5 mostram os resultados do tipo de participantes existentes nas pesquisas: 33,7 % para alunos de Graduação; 25,3% de pesquisadores participantes; 20,6% de pesquisadores coordenadores; 10,2 % de pesquisadores dos cursos *Stricto Sensu*; 7,8 % pesquisadores externos; 1,2 % de pesquisadores dos cursos de *Lato Sensu*; 1,1 % de pesquisadores que são vice-coordenadores.

Os orientadores geralmente trabalham em equipes com alunos de diferentes níveis (graduação, mestrado e doutorado).

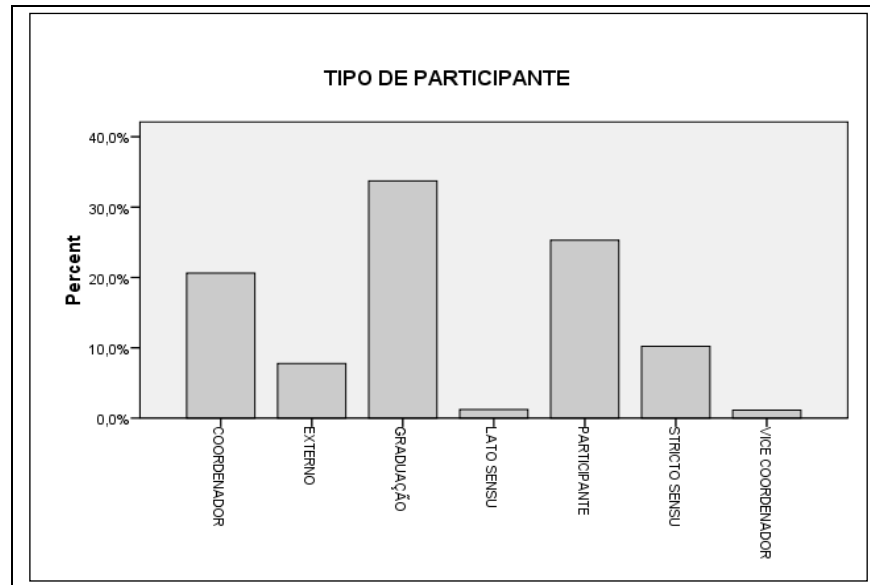


GRÁFICO 5 - Tipo de participante
Fonte: Elaborado pela autora

4.2 ANÁLISE MULTIVARIADA DOS DADOS

Este tópico teve como objetivo fazer uma análise dos dados utilizando a análise multivariada dos dados permitindo um estudo mais amplo.

O *software* utilizado para a análise de agrupamentos foi o SPSS analisando os dados através de cruzamento de categorias escolhidas que foram Formação do orientador, Grandes áreas, e tipo de participante. Através do SPAD 5.5, foram analisando os dados através de clusterização, escolhendo as variáveis como: grandes áreas, palavras-chaves e tipo de participante.

4.2.1 Formação do orientador por grandes áreas

O objetivo foi conhecer qual é o nível de formação do orientador por área. Os resultados do gráfico 6 obtidos pelo *software* mostram que as Ciências Agrárias possuem a maior quantidade de orientadores doutores, com cerca de 24,7%; seguido por Ciências da Saúde com 20,9%; Ciências Humanas com 11,3%; Ciências Biológicas com 10,8%;

Linguística, Letras e Artes com 4,8%; Ciências Exatas e da Terra com 2,1%; e Ciências Sociais Aplicadas possuem 2,1% de doutores.

Nota-se que as áreas que mais investem em pesquisas possuem maior número de doutores e mestres como Ciências da saúde e Ciências Agrárias.

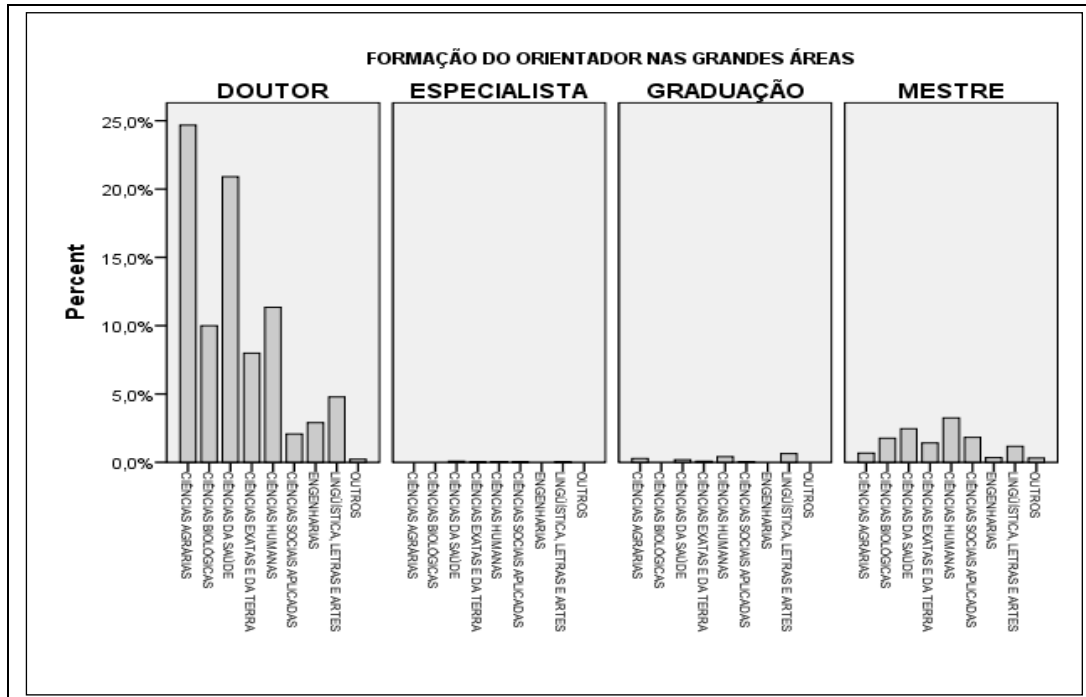


GRÁFICO 6 – Formação do orientador nas grandes áreas

Fonte: Elaborado pela autora

4.2.2 Tipo de participantes por grandes áreas

O gráfico 7 mostra o tipo de participante por grandes áreas. Os resultados mostram que a maioria dos pesquisadores fazem parte dos cursos de Graduação, e pertencem as Ciências Agrárias, 9,7%; Ciências da Saúde, com 7,7%; e Ciências Humanas com 6,6%.

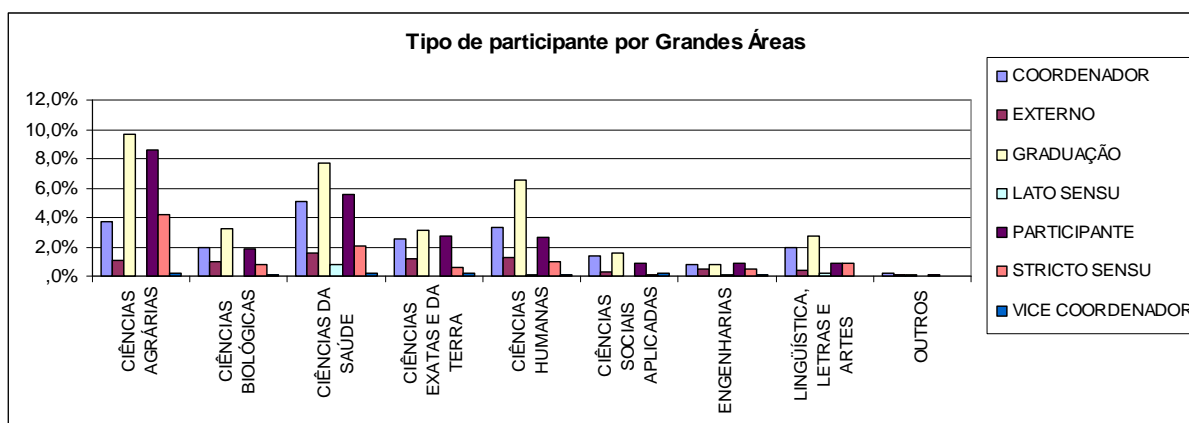


GRAFICO 7 – Tipo de participante por grandes áreas

Fonte: Elaborada pela autora

O fato dos participantes serem em maior parte do curso de Graduação do que os cursos em nível *Lato Sensu* e *Stricto Sensu* pode ser justificado pelo grande incentivo que a universidade oferece aos cursos de graduação como forma de socialização do conhecimento e formação de especialistas, e devido ao grande número de graduandos existentes em cada pesquisa como bolsistas.

4.2.3 Número de palavras -chaves mais freqüentes em relação ao orientador

Em análise realizada pelo *software SPAD*,obteve a lista de palavras mais freqüente em relação ao orientador.

Na figura 02 mostra as listas das palavras mais freqüentes por ordem alfabética e a quantidade de letras das palavras. Verifica-se que as palavras que mais ocorrem são as preposições, como ‘*de*’ (1035 vezes); ‘*e*’ (267 vezes). Embora exista grande número de preposições, o número de palavras mais freqüentes são: cerrado (185 vezes), educação (184 vezes), ensino (180 vezes), saúde (160 vezes) e que podem estar relacionadas a especialidades.

LISTE DES MOTS PAR ORDRE DE FREQUENCE				
NUMERO	MOTS EMPLOYES	FREQUENCES	LONGUEURS	
388	de	1035	2	
467	e	267	1	
451	do	185	2	
260	cerrado	184	7	
473	educacao	184	8	
505	ensino	180	6	
1337	saude	160	5	
1248	qualidade	150	9	
383	da	134	2	
489	em	113	2	
376	cultura	98	7	
500	enfermagem	93	10	
85	analise	90	7	
744	historia	84	8	
79	ambiental	83	9	
229	cancer	81	6	
47	agricultura	79	11	
1504	vida	75	4	
1410	sustentavel	73	11	
266	ciencia	72	7	
1165	plantas	71	7	
351	controle	67	8	
411	desenvolvimento	62	15	
677	geografia	60	9	
1388	solo	60	4	
922	matematica	56	10	
526	erosao	55	6	
336	conservacao	52	11	
630	formacao	50	8	
66	alimentacao	46	11	
52	agua	46	4	

FIGURA 04 – Lista de palavras por ordem de frequência.

Fonte: Pedro Ferri

4.2.4 Segmentos de palavras –chaves

Na figura 05 mostra a lista de segmentos de palavras-chaves por ordem de frequência.

Diferentemente do que foi realizado na figura 04, que obteve os resultados de forma individuais, esta figura mostra as palavras-chaves de forma agrupadas, com base na similaridades. Os dados mostram que, por exemplo, que falou-se muito em: **ensino de** com 99 (noventa e nove) ocorrências; **Qualidade de** com 83 (oitenta e três) ocorrências; **Qualidade de vida** com 55 (cinquenta e cinco) ocorrências; **cerrado** com 52 (cinquenta e dois) ocorrências; **solo** com 48 (quarenta e oito) ocorrências e **Agricultura familiar** 30 (trinta) ocorrências

Observa-se que por ordem de frequência a maior parte das palavras pertencem as grande áreas como Ciências Agrárias, Ciências da Saúde e Educação.

LISTE DES SEGMENTS PAR ORDRE DE FREQUENCE		
NUMERO	FREQUENCE	TEXTE DU SEGMENT
266	99	ensino de
582	83	qualidade de
188	69	de vida
584	55	qualidade de vida
216	52	do cerrado
217	48	do solo
144	44	conservacao de
124	40	ciencia cultura
100	39	cancer de mama
269	39	ensino de geografia
232	38	educacao ambiental
331	36	formacao de
199	35	deterioracao ambiental erosao em pastagem
333	34	formacao de professores
182	33	de plantas
379	33	historia jatai
194	32	desenvolvimento sustentavel
280	32	ensino de ciencia e matematica educacao ambiental
20	31	agricultura sustentavel
19	30	agricultura familiar
175	29	de agua
179	29	de enfermagem
583	28	qualidade de alimentos
37	28	analise do discurso
281	28	ensino de geografia area metropolitana de goiania
178	27	de corte
183	27	de produtos
173	25	da saude
117	25	cerrado bocaiuva
255	24	em saude
464	24	metais pesados

FIGURA 05 – Lista de segmentos de palavras-chave por ordem de frequência
Fonte: Pedro Ferri

4.2.5 Relação entre as palavras-chaves e as grandes áreas

A figura 06 mostra a relação entre as palavras-chaves e as grandes, que foi obtido através da análise multivariada, utilizando o *software* SPAD.

A terceira classe englobou apenas as grandes áreas com as Ciências Biológicas; na segunda classe fazem parte as grandes áreas como as Ciências Exatas e da Terra. As outras áreas se agruparam nas áreas das Ciências Agrárias.

O que se percebe nas *tags* é a similaridade de áreas diferentes que ficaram no mesmo eixo, o que pode ser verificado através das palavras-chaves localizadas nestas áreas. A figura mostra que a palavra **Flora** é mais pesquisada na Classe 2/3, do qual fazem parte as áreas de Ciências exatas e da Terra, e Ciências da saúde, e não na Classe 3/3 formados pelas Ciências Agrárias e Ciências Biológicas. Também pode ser verificado que a grande área da Engenharia não está localizada na classe das Ciências Exatas e da Terra, mas nas Ciências Humanas, fato que pode ser observado através das palavras-chaves '*software*' e '*digital*'.

Na análise realizada na base de dados sobre pesquisa realizada na UFG, verificou-se que vários projetos de pesquisas existentes na grande área de Engenharia, que podem ser relacionados as áreas de Ciências Sociais Aplicadas principalmente na Ciência da informação.

Como exemplo, podem ser citados os projetos como: Programa de proteção da propriedade industrial na escola de engenharia elétrica e de computação realizado na Faculdade de Engenharia Elétrica; e Desenvolvimento de pesquisa e tecnologias de sistemas tutores inteligentes para inclusão digital, que são desenvolvidas na Escola de Engenharia Elétrica e Engenharia da Computação.

Através da mineração de dados, estas palavras-chaves poderiam ser utilizados para verificar as especialidades dos orientadores e as áreas que estes atuam. Os resultados da mineração de dados permitiriam um mapeamento de especialistas através das análises das palavras-chave que são comuns tanto nas Engenharias como nas Ciências Sociais Aplicadas; Lingüística, Letras e Artes; e Ciências Humanas.

As possibilidades e limites de utilização da mineração de dados nos sistemas de cadastro de pesquisa ou no sistema de cadastro de especialistas serão tratados nos seguintes tópicos.

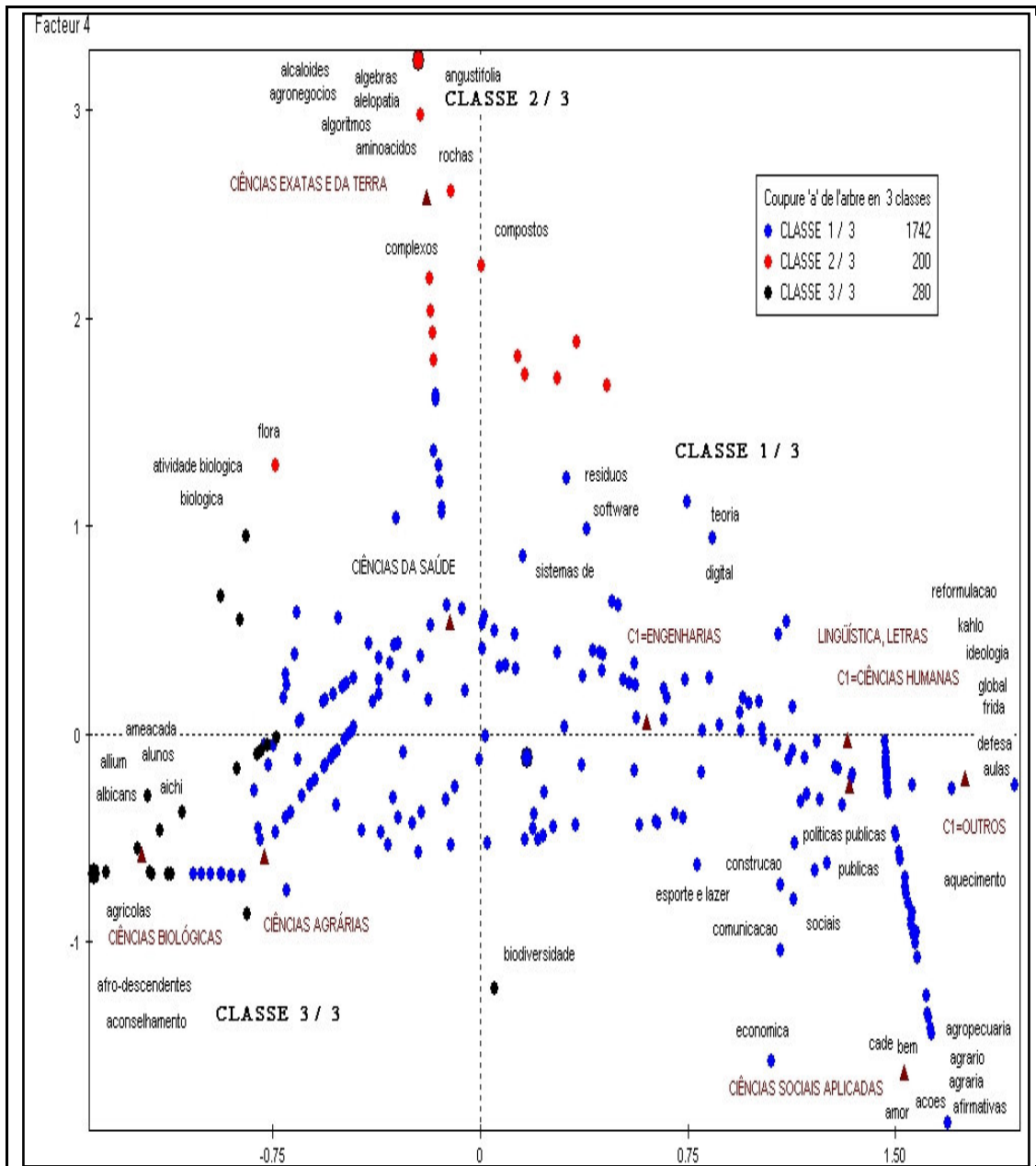


FIGURA 06 - Relação entre as palavras-chaves e as Grandes áreas

Fonte: Pedro Ferri

4.3 APLICABILIDADE DAS TÉCNICAS DE MINERAÇÃO DE DADOS E MINERAÇÃO DE TEXTOS NA UFG: POSSIBILIDADES E LIMITES

A literatura sobre a possibilidade e limites de utilização de mineração de dados para mapear especialistas e especialidades é muito pouco.

A análise ocorreu a partir da análise da literatura encontrada na internet de autores como: Cardoso (2005), com o tema Gestão do conhecimento usando *Data Mining*: estudo de caso na UFLA; Gonçalves (2000), que utilizou de técnicas de mineração de dados em bases

de Ciência & Tecnologia, analisando os grupos de pesquisa no Brasil; Lima (2006), com a dissertação de mestrado abordando a utilização de *data Warehouse* e *Data Mining* no acompanhamento das atividades de pesquisa do Ceulp/Ulbra; e Marcelo de Paula (2004), através da dissertação retratando a exploração da plataforma *Lattes* como fonte de Conhecimento Organizacional em Ciência e Tecnologia.

4.3.1 Possibilidades da mineração de dados

De acordo com autores como Paiva (2007), Paula (2004), Gonçalves (2000), e Cardoso (2005), existem várias possibilidades para a exploração de base de dados sobre cadastro de pesquisas nas IES através das técnicas de mineração de dados.

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

- a) **formação do corpo docente através da identificação de competências e especialidades:** a mineração de dados permitiria um melhor gerenciamento de competências e especialidades de seu corpo docente, através das análises dos dados sobre atividades de pesquisa e extensão cadastrados no banco de dados da instituição. Estes dados seriam obtidos através da comparação entre o nome do orientador e especialidades, facilitando a formação do corpo docente de acordo com a afinidade, formação e atividades correlatas desenvolvidas;
- b) **identificar pesquisas com potencial de geração de patentes:** através da análise quantitativa, permitiria resultados quantitativos de patentes por área de conhecimento; e da análise qualitativa, através da análise de donos das patentes, localização/ país, assunto, domínio, etc.) (OLIVEIRA, 2005);
- c) **identificar linhas de pesquisa em publicações científicas:** a mineração de dados permitiria obter dados estatísticos importantes neste quesito, através da comparação das principais linhas de pesquisa com as publicações de um determinado conjunto de docentes e pesquisadores. Através destes dados, a instituição conheceria as linhas de pesquisas mais relevantes para cada unidade acadêmica;
- d) **identificar interdisciplinaridade e intercessão entre as áreas:** através da mineração de dados, ocorreria o mapeamento das áreas e especialidades diferentes mais que

poderiam contribuir com suas experiências em projetos mais complexos e técnicos. Como exemplo, uma intercessão entre as áreas de Engenharia da Computação, Ciência da Computação com as áreas de Ciência da Informação e Biblioteconomia.

4.3.2 Limites da mineração de dados

Apesar de toda sua capacidade de transformar dados abstratos em informações, o uso de mineração de dados em base de dados sobre grupos de pesquisa científica pode barrar em algumas limitações:

- a) **limites humanos na análise dos dados:** assim como outras metodologias, a utilização das técnicas de mineração de dados deve ser cautelosa, pois muitos resultados emitidos pelos *softwares* são analisados na maioria de forma subjetiva, e muitas vezes uma situação pode ser interpretado como uma regra. Na análise dos dados pode ocorrer uma dificuldade em perceber e interpretar inúmeros fatos observáveis ao longo de processos de mineração (GOLDSCHMIDT; PASSOS, 2005). Berry e Linoff (2000 apud FERNANDES, 2007) atribuem a importância à etapa de entendimento da mineração de dados a comunicação entre as pessoas que entendem o negócio da empresa e pessoas com habilidades quantitativas para transformar dados em conhecimento útil para a tomada de decisão;
- b) ***softwares* e tecnologias:** os *softwares* utilizados para a mineração dos dados devem estar prontos para muitos processos: agregar valor, efetuar conversão, filtrar variáveis, possuir formato de exportação de dados, entre outros. As técnicas de mineração ou mecanismos de busca são usualmente programas ou agentes automatizados inteligentes, incorporando alguma forma de inteligência artificial em bancos de dados relacionais (GOLDSCHMIDT; PASSOS, 2005). Muitas vezes, as técnicas adotadas pelo fornecedor são diferentes e, em consequência, as suas aplicabilidades também não são as mesmas (TAURION, 2008);
- c) **diversidade de formato dos dados:** com base em Braga (2004), pode-se afirmar que em muitos casos, nas instituições como a UFG, as fontes de dados poderão estar em formatos diferentes, como planilhas eletrônicas, bancos de dados. Esta diversidade de

formato dos dados poderá tornar o início da mineração de dados mais trabalhosa, pois, todas as fontes deverão estar reunidas em uma única base de dados ou arquivo, como um *Data Warehouse*. Junta-se a isso, o grande número de dados repetidos, fazendo com que se tenha mais tempo na limpeza dos dados do que na seleção;

- d) **necessidades de testes:** o processo de *Data mining* é cíclico, várias tarefas devem ser executadas alternando-se entre testes de hipóteses e descoberta de conhecimentos (GOLDSCHMIDT; PASSOS, 2005). Embora a ferramenta de mineração de dados seja um importante instrumento para extração de informações úteis perdidas em grandes quantidades de dados aparentemente sem relacionamentos entre si; por outro lado, um uso incorreto desta tecnologia pode gerar muito mais confusão, levando a empresa a tomar decisões erradas (TAURION, 2008);
- e) **escolha de algoritmos:** a dificuldade na escolha de algoritmos de mineração de dados pode gerar resultados insatisfatórios. Os principais produtos de mineração de dados são apenas ferramentas que empregam potentes algoritmos, visando assuntos de negócios e técnicos (GOLDSCHMIDT; PASSOS, 2005);
- f) **necessidade de grande base de dados:** o processo fundamental do *Data Mining* é o de seleção de dados. Sem uma grande base de dados, não será possível a análise que se pretende realizar, e todo o êxito do processo será frustrado (CARVALHO, 2005);
- g) **equipe multidisciplinar:** A mineração de dados é multidisciplinar e origina de diversas áreas, como: Estatística, Inteligência computacional e Aprendizado de máquina, Reconhecimento de padrões, Banco de dados, podendo incluir a Ciência da Informação (GOLDSCHMIDT; PASSOS, 2005). A mineração de dados é uma área de pesquisa que necessita de uma equipe multidisciplinar, incluindo principalmente áreas de 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 dado (CARDOSO, 2005).

O estudo verificou que as ferramentas de mineração de dados ainda não está sendo utilizada de forma bem ampla no contexto das IES, devido ser uma tecnologia nova. De acordo com Carvalho (2005), para ser eficiente, o *Data Mining* precisa ser incorporado à organização e passar a ser uma etapa natural dos outros processos empresariais como o

marketing, as vendas, o suporte ao cliente, o controle de estoque, o planejamento e o projeto do produto.

5 CONSIDERAÇÕES FINAIS

Na Era do conhecimento, tanto as organizações empresariais assim como as IES, o elemento humano tem se constituído numa das peças chave para obter uma competitividade, produtividade e excelência no mercado.

As IES têm construído sistemas de gerenciamento de conhecimento que visam armazenar dados sobre as competências de profissionais. Neste contexto, a mineração de dados pode ser utilizada como uma ferramenta de gestão por competências em IES para identificar características das atividades de pesquisa e de grupos de pesquisas, através do levantamento de perfis, que permitem a seleção de especialistas e orientadores com competências requeridas para a formação de equipes de projeto, no preenchimento de cargos, no planejamento de treinamento, e a verificação da situação das pesquisas além de outras aplicações.

Saber o que ocorrem com os projetos de pesquisas é importante, principalmente quando estas pesquisas são realizadas em parcerias com empresas privadas que muitas vezes concedem apoio financeiro, e em troca, querem por parte das instituições universitárias o direito de prioridade na utilização dos resultados obtidos pelas pesquisas por elas financiadas.

Neste estudo os algoritmos não foram utilizados devido a pequena quantidade de dados e a falta de *softwares* próprios de mineração de dados. Apesar das limitações da pesquisa, as análises dos dados através dos *softwares* utilizados, foi obtido conhecimento importante acerca da funcionalidade da mineração de dados, como os dados podem ser convertidos em conhecimento a partir da análise dos relatórios emitidos pelos *softwares* utilizados e conhecer o perfil das atividades científica da UFG.

Com relação aos resultados apresentados, a conclusão é que estes podem ser utilizados pelos órgãos envolvidos com as coordenações de diversos programas de pesquisa como a PROEC, a PRPPG, a FUNAPE dentre outros. Estes órgãos poderiam aproveitar as informações referentes a baixa quantidade de pesquisas em áreas do conhecimento como Ciência da Informação, Comunicação, Ciências Sociais, Filosofia, Ciência da Computação dentre outros e alocar mais investimentos nestas áreas para a produção de pesquisas.

Os resultados deste estudo não são conclusivos, mas podem ser realizados outros estudos mais eficazes utilizando os algoritmos de mineração de dados. A comparação entre os dados sobre especialidades e o nome do participante não foi possível devido a fraco desempenho de processamento do computador, mas os próximos estudos podem realizar esta tarefa. Outros dados podem ser utilizados para a análise de perfil das atividades científicas da UFG através da mineração de dados, como por exemplo, os dados do Currículo Lattes, pertencente a Plataforma Lattes, e o Diretório de Grupo de pesquisa.

Para a realização de gestão de competências através da mineração de dados, sugere a construção de um sistema inteligente constituída de *softwares* compatível com os *softwares* utilizados pela universidade nos sistemas de informação da UFG. Esta ferramenta permitiria um monitoramento de banco de dados com a tarefa de identificar especialistas e especialidades, para conhecer o perfil dos orientadores e grupos de pesquisas.

A partir do levantamento do perfil, as unidades acadêmicas podem contratar docentes de acordo com a especialidade e o perfil desejado, permitindo um melhor aproveitamento do capital humano existente na instituição e o compartilhamento do conhecimento, e verificar as áreas do conhecimento que possuem *déficit* em relação a atividades de pesquisas.

REFERÊNCIAS

- BARONI, Rodrigo. **Check-up**: avaliando o impacto de seu portal corporativo, parte II. [s.l]: intranetportal, 2005. Disponível em: < <http://www.intranetportal.com.br/tendencias/ti3>>. Acesso em: 6 out. 2008.
- BERNARDES NETO, João. **Tecnologia da informação para o gerenciamento do conhecimento obtido das bases de dados de uma organização**. Dissertação (Mestrado em Engenharia de produção). Departamento de Engenharia de Produção e Sistemas. Universidade Federal de Santa Catarina. Santa Catarina, 2001. Disponível em: < <http://teses.eps.ufsc.br/defesa/pdf/7448.pdf>>. Acesso em: 13 mar. 2008.
- BOTELHO, Maria Antonia; MONTEIRO, Ana Maria; VALLS, Valéria. A gestão do conhecimento esportivo: a experiência da biblioteca da Seme. **Ci. Inf.**, Brasília, v. 36, n. 1, 2007. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S010019652007000100014&lng=&nrm=iso>. Acesso em: 27 out. 2008.
- BRAGA, Luis Paulo Vieira. **Introdução à Mineração de Dados**. 2.ed. Rio de Janeiro : E-papers, 2004.
- BRETZKE, Miriam. Introdução. In: ____ **Marketing de relacionamento e competição em tempo real com CRM (Customer Relationship Management)**. São Paulo: Atlas, 2000. p. 10 – 21.
- _____. CRM transforma a empresa. In: ____ **Marketing de relacionamento e competição em tempo real com CRM (Customer Relationship Management)**. São Paulo: Atlas, 2000. p. 28 – 30.
- _____. Oportunidades de negócio com o *Data Mining*. In: ____ **Marketing de relacionamento e competição em tempo real com CRM (Customer Relationship Management)**. São Paulo: Atlas, 2000. p. 160-169 .
- CAMISASCA, Flávia. UFMG prepara catálogo eletrônico de especialistas e especialidades. **Boletim informativo**. nº 1498, ano 31, 2005. Disponível em: < <http://www.ufmg.br/boletim/bol1498/quarta.shtml>>. Acesso em: 6 out. 2008.
- CAMPOS, Pedro. **Sessões práticas com software**. Porto: Faculdade de Economia. Universidade do Porto, 2003. Disponível em: < <http://www.fep.up.pt/disciplinas/ce707/software.htm>>. Acesso em:
- CARDOSO, Olinda Nogueira Paes. Gestão do conhecimento usando o Data Mining: estudo de caso na UFLA. 2005. Dissertação (Mestrado em Administração). Universidade Federal de Lavras. Minas Gerais, 2005. Disponível em: < http://bibtede.ufla.br/tede//tde_busca/arquivo.php?codArquivo=255>. Acesso em: 13 set. 2008.
- CARDOSO, Gilson da Silva, FRANCISCO, Antonio Carlos e SCANDELARI, Luciano. Banco de competências – proposta de uma ferramenta de apoio à decisão de capacitação de Rh. In: ____ PILATTI, Luiz Alberto (Org.). **Temas em engenharia de produção**.

Jundiaí: Editora Fontoura, 2005. Disponível em: <http://www.pg.cefetpr.br/ppgep/livro/4_Gilson_Cardoso.pdf>. Acesso em: 21 ago. 2008.

CARVALHO, Luis Alfredo Vidal de. **Data Mining: A mineração de dados no marketing, Medicina, Economia, Engenharia e Administração.** Rio de Janeiro: Editora Ciência Moderna, 2005. p. 9 - 143.

CHERUBINI NETO, Reinaldo. **A criação do conhecimento na relação Universidade-empresa:** estudo do caso de desenvolvimento de um *software* de configuração para um equipamento de rede. Dissertação (Mestrado em Administração) - Escola de Administração. Universidade Federal do Rio Grande do Sul. Rio Grande do Sul, 2003. Disponível em: <http://volpi.ea.ufrgs.br/teses_e_dissertacoes/td/000969.pdf>. Acesso em: 08 nov. 2008

CHOO, C. W. Como ficamos sabendo. In: _____ **A organização do conhecimento:** como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões. São Paulo: SENAC, 2003. p. 65-120.

_____ A administração da ambigüidade. In: _____ **A organização do conhecimento:** como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões. São Paulo: SENAC, 2003. p. 121-251.

CONNECTT. **Mapeando conhecimento com páginas amarelas.** Porto Alegre: Conectt, 2006. Disponível em: <http://www.conectt.com.br/objs/Noticias/editorial_news_junho2006.asp>. Acesso em: 06 set. 2008.

CORRÊA, Ulisses. **Mineração de dados de help desk usando rattle:** o caso Petrobras. Dissertação (Mestrado em Administração). 2007. Faculdade de Economia e Finanças da IBMEC. Rio de Janeiro, 2007. Disponível em: <http://www.ibmecrj.br/sub/RJ/files/dissert_mestrado/ADM_ulissescorrea_fev.pdf>. Acesso em: 30 out. 2008.

CÔRTEZ, S da Costa; PORCARO Rosa Maria; LIFSCHITZ, Sérgio. **Mineração de Dados:** Funcionalidades, técnicas e abordagens. Rio de Janeiro: PUC, 2002. Disponível em: <ftp://ftp.inf.puc-rio.br/pub/docs/techreports/02_10_cortes.pdf>. Acesso em: 10 abr. 2008.

DAVENPORT, Thomas H. A informação e seus dissabores. In: _____ **Ecologia da Informação.** São Paulo: Editora Futura, 2001. p. 18.

FANDERUFF, Damaris; MARTINS, Alejandro. Solução de DW para análise estratégica dos Indicadores de desempenho (ids) do estado de SC. In: _____ ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 22, 2002, Curitiba. **Anais eletrônicos...** Curitiba, 2002. Disponível em: <http://www.abepro.org.br/biblioteca/ENEGEP2002_TR92_1235.pdf>. Acesso em: 30 out. 2008.

FERNANDES, Marcelo Pires. **Descoberta do conhecimento em base de dados e estratégias de relacionamento com clientes:** um estudo no setor de serviços. Dissertação (Mestrado em Administração de empresa) – Faculdade de Ciências Econômicas, Instituto Presbiteriano Mackenzie. São Paulo, 2007. Disponível em: <http://mx.mackenzie.com.br/tde/tde_arquivos/1/TDE-2008-02-05T130440Z-381/Publico/Marcelo%20Pires%20Fernandes.pdf>. Acesso em: 23 out. 2008.

FLEURY, Afonso; FLEURY, Maria Tereza Leme. **Estratégias empresariais e formação de competências**: um quebra-cabeça caleidoscópico da indústria brasileira. 3 ed. São Paulo: Atlas, 2006. p. 26 – 82.

FREITAS JUNIOR, Olival de Gusmão. **Um modelo de sistema de gestão do conhecimento para grupos de pesquisa e desenvolvimento**. Tese (Doutorado em Engenharia de Produção). Universidade Federal de Santa Catarina. Florianópolis, 2003. Disponível em: <teses.eps.ufsc.br/defesa/pdf/9835.pdf>. Acesso em: 23 out. 2008.

GOLDSCHMIDT, Ronaldo; PASSO, Emmanuel. **Data Mining**: um guia prático. Rio de Janeiro: Elsevier, 2005. p.1- p.116.

GONÇALVES, A. L.. **Utilização de Técnicas de Mineração de Dados em Bases de C&T**: uma Análise dos Grupos de Pesquisa no Brasil. Dissertação (Mestrado em Engenharia de produção). Universidade Federal de Santa Catarina. Florianópolis, 2000. Disponível em: <<http://teses.eps.ufsc.br/defesa/pdf/3735.pdf>>. Acesso em: 23 out. 2008.

JESUS, Alberto Pereira de. **Data mining aplicado à identificação do perfil dos usuários de uma biblioteca para a personalização de sistemas web de recuperação e disseminação de informações**. 2004. 120 f. Dissertação (Mestrado em Ciência da Computação). Universidade Federal de Santa Catarina. Florianópolis, 2004. 120p. Disponível em: <<http://150.162.90.250/teses/PGCC0587.pdf>>. Acesso em: 12 fev. 2008.

LIMA, Helena Carolina Costa e. **Utilização de Data Warehouse e Data Mining no acompanhamento das atividades de pesquisa do Ceulp/Ulbra**. Trabalho de conclusão de curso (Bacharelado de sistema de informação). Universidade Luterana do Brasil. Palmas, 2006. Disponível em:<<http://www.ulbra-to.br/ensino/43020/artigos/relatorios20061/Arquivos/Helena%20C%20L%20Trabalho%20de%20Conclusao%20de%20Curso.pdf>> . Acesso em: 14 mar. 2008.

LEITE, Fernando Cesar Lima de. **Gestão do conhecimento científico no contexto acadêmico**: proposta de um modelo conceitual. Dissertação (Mestrado em Ciência da Informação) - Departamento de Ciência da Informação. Universidade de Brasília. Distrito Federal, 2006. Disponível em: <http://eprints.rclis.org/archive/00006259/01/Gest%C3%A3o_do_conhecimento_cient%C3%ADfico_no_contexto_acad%C3%AAmico_-_Fernando_2006.pdf>. Acesso em: 16 out. 2008.

LOJKINE, Jean. **A revolução informacional**. São Paulo: Cortez, 1995. p. 49 – 109.

MALHOTRA, Naresh K. **Concepção de pesquisa**. In ____: Pesquisa em marketing: uma orientação aplicada. Bookman, 2006. p.98-120.

MORAES, Rachel Rubia Vellani Bueno de; GOMES, Rodrigo Cavalcanti de Brito; FRANKEN, Tjerk Guus. **Mapeamento das competências da ESPM Corporate**. Trabalho de conclusão de curso (Especialização em Engenharia da Produção) – COPPE. Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2005. Disponível em: <portal.crie.coppe.ufrj.br/.../%7B7CA9E523-AD8A-45C1-88D6-D01BA691BEA8%7D/RJ11_Projeto_03.pdf>. Acesso em: 29 ago. 2008.

NASCIMENTO, Marcelo Magno Rocha. Gestão do conhecimento: desafio ou necessidade. **Administradores**. Vitória: administradores, 2006. Disponível em: < http://www.administradores.com.br/producao_academica/gestao_do_conhecimento_desafio_ou_necessidade/160/>. Acesso em: 31 ago. 2008.

NICHOLSON, Scoth. O processo de Bibliomineração: repositório de dados para tomada de decisão em bibliotecas. Trad. Prof. Dr. Ulf Gregor Baranow. **Transinformação**. Campinas, v.16, n.3, p.253-261, set/dez, 2004. Disponível em: <<http://revistas.puc-campinas.edu.br/transinfo/viewarticle.php?id=74>>. Acesso em: 29 de Maio de 2008.

OLIVEIRA, Jonice. **METHEXIS**: um ambiente de apoio à gestão do conhecimento em “e-science. [slides]. Rio de Janeiro: UFRJ/COPPE, 2005. Disponível em: < www.cos.ufrj.br/~jonice/qualificacao/Qualificacao_JoniceOliveira.ppt >. Acesso em: 01 nov. 2008.

PAIVA, Luiz Henrique de. **Informação não é nada sem conhecimento**. São Paulo, 2007. Disponível em: <<http://ogereente.com/empreendaja/2007/09/17/informacao-nao-e-nada-sem-conhecimento/>>. Acesso em 6 fev. 2008.

PAULA, Marcelo V. de. **Explorando o Potencial da Plataforma Lattes como Fonte de Conhecimento Organizacional em Ciência e Tecnologia**. Dissertação (Mestrado). Brasília: UCB, 2004. Disponível em: < http://www.bdttd.ucb.br/tede/tde_busca/arquivo.php?códArquivo=160 >. Acesso em: 29 set. 2008.

PENTEADO, Roberto. Análise e mineração de dados e textos. In: _____DUARTE, Jorge; BARROS, Antonio (Org.). **Métodos e técnicas de pesquisa em comunicação**. 2. ed. São Paulo: Atlas: 2006. Disponível em: < http://www.editoraatlas.com.br/Atlas/portal/ProductDetail.ctrl.aspx?product_id=8522445338>. Acesso em: 11 de Jun. de 2008.

PICANÇO, Caroline Tavares Picanço. **Um estudo sobre a gestão da competência adotada em uma *software house***. Trabalho de conclusão de curso (Graduação em Tecnologia em Desenvolvimento de *Software*). Centro Federal de Educação Tecnológica do Amazonas. Manaus, 2006. Disponível em: < www.scribd.com/doc/2348778/Um-estudo-de-caso-sobre-a-gestao-por-competencia-adotada-em-uma-software-house>. Acesso em: 24 ago. 2008

PINHEIRO, Carlos André Reis. **Web Warehousing**: extração e gerenciamento de dados na internet. Rio de Janeiro: Axcel Brooks, 2003.

PLATAFORMA LATTES. Conteúdo da Plataforma Lattes . Distrito Federal: Plataforma lattes, 2008. Disponível em: < <http://lattes.cnpq.br/conteudo/estatisticas.htm> >. Acesso em: 08 nov. 2008.

QUONIAM, Luc et al . Inteligência obtida pela aplicação de *data mining* em base de teses francesas. **Ciência da Informação**, Brasília, v. 30, n. 2, 2001. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652001000200004&lng=en&nrm=iso>. Acesso em: 30 mar. 2008.

REZENDE, Denis Alcides. Sistemas de conhecimento e as relações com a gestão do conhecimento e com a inteligência organizacional nas empresas privadas e nas organizações públicas. In: _____ TARAPANOFF, Kira (Org.). **Inteligência, Informação e conhecimento em corporações**. Brasília: IBICT/ UNESCO, 2006. p.257-277.

RUIZ, João Álvaro. **Metodologia Científica**: guia para eficiência nos estudos. 5. ed. São Paulo: Atlas, 2002.

SANTOS, Elisa Gomes dos; CERANTE, Livia Luz. **Gestão do Conhecimento**: um estudo para facilitar sua implantação nas empresas. Trabalho de conclusão de curso (Curso de Informática) – Instituto de Matemática. Universidade Federal do Rio de Janeiro. Rio de Janeiro: UFRJ, 2000. Disponível em: < <http://dataware.nce.ufrj.br:8080/dataware/publicacoes/dataware/fisico/projetosFinais/gestaodoconhecimento/SANTOS-CERANTE-2000.pdf> >. Acesso em: 08 nov. 2008.

SETZER, Valdemar. Como mapear competências?. **Revista Programação**. São Paulo, ano 4, n.29, p. 13 e 18, 2001. Disponível em: < <http://www.ime.usp.br/~vwsetzer/comp-Sucesu.html> >. Acesso em: 24 ago. 2008.

SILVA, Adriane Carla Anastácio da; FRANCISCO, Antônio Carlos de e SCANDELARI, Luciano. A tecnologia da informação aplicada ao mapeamento de competências: Caso Cia Iguazu de Café Solúvel. In: _____ ENCONTRO DE ENGENHARIA ,E TECNOLOGIA DOS CAMPOS GERAIS, 3, 2007, Ponta Grossa (PR). **Anais eletrônicos...** Ponta Grossa, 2007. Disponível em: < <http://www.aeapg.org.br/3eetcg/Anais/ARTIGOS/PDFS/Engenharia%20de%20Produ%E3o%20-%202005.pdf> >. Acesso em: 21 ago. 2008.

SILVA, Fábio Mascarenhas e. **Organização da informação em sistemas eletrônicos abertos de Informação Científica & Tecnológica**: Análise da Plataforma Lattes. Tese (Doutorado). São Paulo: USP, 2007. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/27/27151/tde-17032008-095556/>>. Acesso em: 26 de Set. de 2008.

SILVA FILHO, Antônio Mendes da. **Gestão do Conhecimento: sobre a Importância da Extração da Informação**. **Revista espaço acadêmico**. n.59, Universidade Estadual de Maringá. Maringá, 2006. Disponível em: < <http://www.espacoacademico.com.br/059/59silvafilho.htm> >. Acesso em: 06 nov. 2008.

SOUZA, Michel de. **Data mining**. Vitória, ES: Imasters online, 2003. Disponível em: http://imasters.uol.com.br/artigo/1482/bi/data_mining/. Acesso em: 13 abr. 2008.

STRAUHS, Faimara do Rocio; ABREU, Aline França; RENAUX, Douglas P. B. O mapeamento de competências como ferramenta auxiliar do processo de gestão do conhecimento. In: _____ SEMINARIO LATINO-IBEROAMERICANO DE GESTIÓN TECNOLÓGICA/ALTEC, 9, 2001, Costa Rica. **Anais eletrônicos...** Costa Rica, 2001. Disponível em: < www.lit.cpdtt.cefetpr.br/P_D/biblio/mapa_altec.pdf >. Acesso em: 23 out. 2008.

TAURION, Cezar. **Qual a definição para data mining?**. Blumenau, SC: Marketing direto, 2007. Disponível em: <http://www.mktdireto.com.br/materia6.html>. Acesso em: 13 abr. 2008.

TURBAN, Efraim; RAINER JUNIOR, R. Kelly; POTTER, Richard E. Gerenciamento de dados e conhecimento. In: _____ **Administração de tecnologia da informação**: teoria e prática. Rio de Janeiro: Elsevier, 2005. p. 74-113.

VAITSMAN, Hélio Santiago. Informação como sobrevivência. In: ____ **Inteligência empresarial: atacando e defendendo**. Rio de Janeiro: Editora Interciência, 2001. p.18.

WEITZEN, H. Skip. **O poder da informação: como transformar a informação que você domina em um negócio lucrativo**. São Paulo: Makron, McGraw – Hill, 1991.

ANEXOS

ANEXO A – Frequência dos resultados extraído com o *software* SPSS Statistics 17.0 sobre os dados da pesquisa científica da UFG

	GRANDE AREA	Frequency	Percent
	OUTROS	31	,5
	LINGÜÍSTICA, LETRAS E ARTES	384	6,6
	ENGENHARIAS	188	3,3
	CIÊNCIAS SOCIAIS APLICADAS	230	4,0
	CIÊNCIAS HUMANAS	871	15,1
	CIÊNCIAS EXATAS E DA TERRA	550	9,5
	CIÊNCIAS DA SAÚDE	1367	23,6
	CIÊNCIAS BIOLÓGICAS	680	11,8
	CIÊNCIAS AGRÁRIAS	1482	25,6
	Total	5783	100,0

	ÁREA DO CONHECIMENTO	Frequency	Percent
	AGRONOMIA	590	10,2
	MEDICINA	374	6,5
	MEDICINA VETERINÁRIA	363	6,3
	EDUCAÇÃO	303	5,2
	ZOOTECNIA	271	4,7
	ODONTOLOGIA	214	3,7
	CIÊNCIA E TECNOLOGIA DE ALIMENTOS	204	3,5
	SAÚDE COLETIVA	204	3,5
	FARMÁCIA	198	3,4
	LINGÜÍSTICA E LETRAS	189	3,3
	HISTÓRIA	186	3,2
	ENFERMAGEM	180	3,1
	QUÍMICA	160	2,8
	MICROBIOLOGIA	156	2,7
	FÍSICA	133	2,3

GEOGRAFIA	132	2,3
ARTES	121	2,1
NUTRIÇÃO	116	2,0
PSICOLOGIA	107	1,9
ENGENHARIA CIVIL	102	1,8
MATEMÁTICA	91	1,6
GENÉTICA	89	1,5
GEOCIÊNCIAS	89	1,5
DIREITO	84	1,5
LETRAS	74	1,3
MORFOLOGIA	73	1,3
CIÊNCIA DA COMPUTAÇÃO	71	1,2
SOCIOLOGIA	70	1,2
ADMINISTRAÇÃO	69	1,2
ECOLOGIA	62	1,1
FARMACOLOGIA	53	,9
BIOQUÍMICA	48	,8
ENGENHARIA ELÉTRICA	48	,8
	47	,8
BOTÂNICA	46	,8
EDUCAÇÃO FÍSICA	43	,7
ZOOLOGIA	41	,7
ECONOMIA	30	,5
FISIOLOGIA	30	,5
CIÊNCIA POLÍTICA	29	,5
COMUNICAÇÃO	29	,5
PARASITOLOGIA	29	,5
IMUNOLOGIA	28	,5
MULTIDISCIPLINAR	26	,4
ENGENHARIA AGRÍCOLA	24	,4
FILOSOFIA	22	,4
BIOLOGIA GERAL	21	,4
ENGENHARIA SANITÁRIA	19	,3
ARQUEOLOGIA	16	,3
FISIOTERAPIA E TERAPIA OCUPACIONAL	13	,2
RECURSOS PESQUEIROS E ENGENHARIA DE PESCA	13	,2
ANTROPOLOGIA	8	,1

	CIÊNCIA DA INFORMAÇÃO	8	,1
	ENGENHARIA BIOMÉDICA	7	,1
	ENGENHARIA DE PRODUÇÃO	5	,1
	ENGENHARIA NAVAL E OCEÂNICA	5	,1
	TURISMO	5	,1
	ENSINO	3	,1
	PLANEJAMENTO URBANO E REGIONAL	3	,1
	PROBABILIDADE E ESTATÍSTICA	3	,1
	DEMOGRAFIA	2	,0
	ENGENHARIA MECÂNICA	2	,0
	RECURSOS FLORESTAIS E ENGENHARIA FLORESTAL	2	,0
	Total	5783	100,0

	FORMAÇÃO DO ORIENTADOR	Frequency	Percent
	DOUTOR	4910	84,9
	MESTRE	765	13,2
	GRADUAÇÃO	96	1,7
	ESPECIALISTA	12	,2
	Total	5783	100,0

	ORGAO DO ORIENTADOR	Frequency	Percent
	CAMPUS JATAÍ	853	14,8
	ESCOLA DE AGRONOMIA E ENGENHARIA DE ALIMENTOS	610	10,5
	ESCOLA DE VETERINÁRIA	600	10,4
	CAMPUS CATALÃO	427	7,4
	INSTITUTO DE CIÊNCIAS BIOLÓGICAS	390	6,7
	INSTITUTO DE PATOLOGIA TROPICAL E SAÚDE PÚBLICA	340	5,9
	FACULDADE DE FARMÁCIA	270	4,7
	FACULDADE DE ODONTOLOGIA	270	4,7
	FACULDADE DE ENFERMAGEM	232	4,0

FACULDADE DE NUTRIÇÃO	201	3,5
FACULDADE DE MEDICINA	196	3,4
FACULDADE DE CIÊNCIAS HUMANAS E FILOSOFIA	159	2,7
FACULDADE DE LETRAS	155	2,7
ESCOLA DE ENGENHARIA CIVIL	105	1,8
INSTITUTO DE QUÍMICA	102	1,8
INSTITUTO DE ESTUDOS SÓCIO-AMBIENTAIS	98	1,7
ESCOLA DE MÚSICA E ARTES CÊNICAS	97	1,7
CENTRO DE ENSINO E PESQUISA APLICADA À EDUCAÇÃO	89	1,5
FACULDADE DE EDUCAÇÃO	89	1,5
INSTITUTO DE FÍSICA	87	1,5
FACULDADE DE EDUCAÇÃO FÍSICA	65	1,1
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA	58	1,0
FACULDADE DE DIREITO	53	,9
ESCOLA DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO	51	,9
FACULDADE DE COMUNICAÇÃO E BIBLIOTECONOMIA	37	,6
FACULDADE DE DIREITO - EXTENSÃO GOIÁS	37	,6
HOSPITAL DAS CLÍNICAS	35	,6
INSTITUTO DE INFORMÁTICA	35	,6
FACULDADE DE ARTES VISUAIS	22	,4
UFG	8	,1
DEPARTAMENTO DE ASSUNTOS ACADÊMICOS	7	,1
PRÓ-REITORIA DE ASSUNTOS DA COMUNIDADE UNIVERSITÁRIA	5	,1
Total	5783	100,0

	TIPO DE PARTICIPANTE	Frequency	Percent
	GRADUAÇÃO	1950	33,7
	PARTICIPANTE	1463	25,3
	COORDENADOR	1193	20,6
	STRICTO SENSU	591	10,2
	EXTERNO	449	7,8
	LATO SENSU	71	1,2
	VICE COORDENADOR	66	1,1
	Total	5783	100,0

	TIPO DE PARTICIPANTE POR GRANDE ÁREA						
	COORDENADOR	EXTERNO	GRADUADO	LATO SENSU	PARTICIPANTE	STRICTO SENSU	VICE COORDENADOR
CIÊNCIAS AGRÁRIAS	199	58	514	0	453	220	9
	3,8%	1,1%	9,7%	,0%	8,5%	4,2%	,2%
CIÊNCIAS BIOLÓGICAS	103	50	171	0	99	39	5
	1,9%	,9%	3,2%	,0%	1,9%	,7%	,1%
CIÊNCIAS DA SAÚDE	270	82	406	43	297	110	12
	5,1%	1,5%	7,7%	,8%	5,6%	2,1%	,2%
CIÊNCIAS EXATAS E DA TERRA	132	63	168	0	147	32	8
	2,5%	1,2%	3,2%	,0%	2,8%	,6%	,2%
CIÊNCIAS HUMANAS	176	66	348	4	139	54	5
	3,3%	1,2%	6,6%	,1%	2,6%	1,0%	,1%
CIÊNCIAS SOCIAIS APLICADAS	71	16	81	0	47	7	8
	1,3%	,3%	1,5%	,0%	,9%	,1%	,2%
ENGENHARIAS	40	26	39	4	46	28	5
	,8%	,5%	,7%	,1%	,9%	,5%	,1%
LINGÜÍSTICA, LETRAS E ARTES	104	19	143	9	45	47	1
	2,0%	,4%	2,7%	,2%	,8%	,9%	,0%
OUTROS	12	6	6	0	4	2	1
	,2%	,1%	,1%	,0%	,1%	,0%	,0%
Total	1107	386	1876	60	1277	539	54
	20,9%	7,3%	35,4%	1,1%	24,1%	10,2%	1,0%

ANEXO B – Figuras de formatação dos dados do *software* SPAD 5.5

Editeur de données - ke-4525 casos

Fichier Edition Affichage Fenêtre ?

	Idem	Libl	GRAN	AREA	SUBA	ESPE	ORIE	FORM	ORGA	NOME	TIPO	Q001	Q002
1	I0001	0001	5	18			249	1	13	493	3	educação infantil	políticas públicas
2	I0002	0002	5	18			249	1	13	964	3	educação infantil	políticas públicas
3	I0003	0003	5	18			249	1	13	2833	3	educação infantil	políticas públicas
4	I0004	0004	5	18			249	1	13	2700	3	educação infantil	políticas públicas
5	I0005	0005	5	18			249	1	13	1273	1	educação infantil	políticas públicas
6	I0006	0006	8	42	141		390	1	2	2064	1		
7	I0007	0007	8	42	141		390	1	2	1304	3		
8	I0008	0008	8	42	205		28	2	1	132	1	sociolinguística	cultura popular /religiosidade
9	I0009	0009	8	42	205		28	2	1	1967	2	sociolinguística	cultura popular /religiosidade
10	I0010	0010	3	58	200		121	1	19	2952	3	PROMOÇÃO DA SAÚDE	SAÚDE BUCAL COLETIVA
11	I0011	0011	3	58	200		121	1	19	615	1	PROMOÇÃO DA SAÚDE	SAÚDE BUCAL COLETIVA
12	I0012	0012	3	58	200		121	1	19	22	3	PROMOÇÃO DA SAÚDE	SAÚDE BUCAL COLETIVA
13	I0013	0013	3	58	200		121	1	19	2363	3	PROMOÇÃO DA SAÚDE	SAÚDE BUCAL COLETIVA
14	I0014	0014	3	19	64		487	1	18	408	4		
15	I0015	0015	3	19	64		487	1	18	2634	1		
16	I0016	0016	3	19	64		487	1	18	2895	4		
17	I0017	0017	3	58	83	99	41	1	26	2643	3	Staphylococcus aureus	epidemiologia molecular
18	I0018	0018	3	58	83	99	41	1	26	3266	6	Staphylococcus aureus	epidemiologia molecular
19	I0019	0019	3	58	83	99	41	1	26	2872	3	Staphylococcus aureus	epidemiologia molecular
20	I0020	0020	3	58	83	99	41	1	26	2836	5	Staphylococcus aureus	epidemiologia molecular
21	I0021	0021	3	58	83	99	41	1	26	2706	5	Staphylococcus aureus	epidemiologia molecular
22	I0022	0022	3	58	83	99	41	1	26	2044	5	Staphylococcus aureus	epidemiologia molecular
23	I0023	0023	3	58	83	99	41	1	26	2035	6	Staphylococcus aureus	epidemiologia molecular
24	I0024	0024	3	58	83	99	41	1	26	1891	6	Staphylococcus aureus	epidemiologia molecular
25	I0025	0025	3	58	83	99	41	1	26	1468	6	Staphylococcus aureus	epidemiologia molecular
26	I0026	0026	3	58	83	99	41	1	26	252	5	Staphylococcus aureus	epidemiologia molecular
27	I0027	0027	3	58	83	99	41	1	26	211	1	Staphylococcus aureus	epidemiologia molecular
28	I0028	0028	5	12	182		570	4	29	528	5	patrimônio cultural imaterial	conservação de acervos

Tela de edição com dados que foram codificados.

Elaborado pelo professor Pedro Ferri com o *software* SPAD 5.5.

CARACTERISATION PAR LES FREQUENCES DES CLASSES OU MODALITES							
DE Coupure 'a' de l'arbre en 3 classes							
CLASSE 1 / 3							
V.TEST	PROBA	POURCENTAGES			FREQUENCES CARACTERISTIQUES		POIDS
		CLA/FRE	FRE/CLA	GLOBAL	NUM .	LIBELLE	
				85.46		CLASSE 1 / 3	aa1a 18095
40.52	0.0000	99.25	29.95	25.79	1 .	CIÊNCIAS AGRÁRIAS	AA_1 5461
38.96	0.0000	98.61	31.32	27.15	3 .	CIÊNCIAS DA SAÚDE	AA_3 5748
30.63	0.0000	99.64	16.97	14.55	5 .	CIÊNCIAS HUMANAS	AA_5 3081
22.08	0.0000	100.00	8.37	7.16	8 .	LINGÜÍSTICA, LETRAS	AA_8 1515
14.00	0.0000	98.92	4.55	3.93	7 .	CIÊNCIAS ENGENHARIAS	AA_7 833
12.52	0.0000	98.04	4.42	3.85	6 .	CIÊNCIAS SOCIAIS APL	AA_6 815
3.54	0.0002	100.00	0.30	0.26	9 .	CI=OUTROS	AA_9 54
				5.12		CLASSE 2 / 3	aa2a 1083
78.46	0.0000	74.10	94.55	6.53	4 .	CIÊNCIAS EXATAS E DA	AA_4 1382
				9.42		CLASSE 3 / 3	aa3a 1995
96.72	0.0000	81.22	92.98	10.79	2 .	CIÊNCIAS BIOLÓGICAS	AA_2 2284

Classes que representam as Grandes Áreas na matriz multivariada.

Elaborado pelo professor Pedro Ferri com o software SPAD 5.5