

ANDRÉ LUIZ SALES

**ANÁLISE DE REFERÊNCIAS EM TESES COMO APOIO À TOMADA DE
DECISÃO NO DESENVOLVIMENTO DE COLEÇÕES DA BIBLIOTECA
UNIVERSITÁRIA**

Monografia de graduação apresentada ao Departamento de Ciência da Computação da Universidade Federal de Lavras como parte das exigências do curso de Ciência da Computação para obtenção do título de Bacharel em Ciência da Computação.

LAVRAS
MINAS GERAIS - BRASIL
2007

ANDRÉ LUIZ SALES

**ANÁLISE DE REFERÊNCIAS EM TESES COMO APOIO À TOMADA DE
DECISÃO NO DESENVOLVIMENTO DE COLEÇÕES DA BIBLIOTECA
UNIVERSITÁRIA**

Monografia de graduação apresentada ao Departamento de
Ciência da Computação da Universidade Federal de
Lavras como parte das exigências do curso de Ciência da
Computação para obtenção do título de Bacharel em
Ciência da Computação.

Área de Concentração:
Banco de Dados

Orientadora:
Prof^a Olinda Nogueira Paes Cardoso

LAVRAS
MINAS GERAIS - BRASIL
2007

**Ficha Catalográfica preparada pela Divisão de Processos Técnico
da Biblioteca Central da UFLA**

Sales, André Luiz

Análise de Citações de Teses como Apoio à Tomada de Decisão no Desenvolvimento de Coleções da Biblioteca Universitária/ André Luiz Sales. Lavras – Minas Gerais, 2007. 48. p : il.

Monografia de Graduação –Universidade Federal de Lavras. Departamento de Ciência da Computação.

1. Banco de Dados. 2. Ciência da Informação. 3. Bibliometria. I. SALES, A. L. II. Universidade Federal de Lavras. III. Análise de Citações de Teses como Apoio à Tomada de Decisão no Desenvolvimento de Coleções da Biblioteca Universitária.

ANDRÉ LUIZ SALES

**ANÁLISE DE REFERÊNCIAS EM TESES COMO APOIO À TOMADA DE
DECISÃO NO DESENVOLVIMENTO DE COLEÇÕES DA BIBLIOTECA
UNIVERSITÁRIA**

Monografia de graduação apresentada ao Departamento de
Ciência da Computação da Universidade Federal de
Lavras como parte das exigências do curso de Ciência da
Computação para obtenção do título de Bacharel em
Ciência da Computação.

Aprovada em 10 de agosto de 2007

Professor André Luiz Zambalde

Professor André Vital Saúde

Prof^a. Olinda Nogueira Paes Cardoso
(Orientadora)

LAVRAS
MINAS GERAIS – BRASIL

Agradecimentos

Agradeço a Deus em primeiro lugar por ter me dado perseverança e força para não desistir em atingir meu objetivo de finalizar o curso. Agradeço a meus pais e minhas irmãs por me apoiarem em todos os momentos e por serem meu porto seguro. Aos amigos agradeço pela jornada compartilhada e pelos bons momentos vividos, e agradeço especialmente a professora Olinda Nogueira Paes Cardoso, que além dos ensinamentos técnicos me passou ensinamentos de vida e me concedeu a honra de sua amizade.

Análise de referências em teses como apoio à tomada de decisão no desenvolvimento de coleções da biblioteca universitária

RESUMO

O trabalho aqui apresentado realiza o estudo do uso da informação técnica e científica, por meio da análise das referências utilizadas na elaboração de teses acadêmicas. A pesquisa quantitativa usa a bibliometria para apoiar a tomada de decisão no desenvolvimento de coleções de uma biblioteca. Analisa-se o conjunto de dados extraídos das teses defendidas entre os anos de 1999 a 2006, orientadas por professores pertencentes ao Departamento de Ciências Administrativas da Escola de Administração da Universidade Federal do Rio Grande do Sul (UFRGS) representando, até esta data, um total de 80 teses disponibilizadas em texto completo na Internet. O objetivo geral é criar um banco de dados das referências bibliográficas de teses de um programa de pós-graduação e posteriormente estudar o uso da informação técnica e científica por meio da análise das citações. Os objetivos específicos são coletar as teses disponíveis em formato digital, que irão servir de base para a análise; modelar e desenvolver um banco de dados das referências bibliográficas utilizadas em todas as teses; e criar consultas no banco de dados permitindo analisar a tipologia dos materiais bibliográficos utilizados nas citações, sua obsolescência, os títulos de periódicos citados e o idioma dos materiais consultados. Os resultados obtidos demonstram que as informações são úteis para a tomada de decisão dos administradores da biblioteca, no que diz respeito ao desenvolvimento de suas coleções.

Palavras-chave: Banco de Dados. Análise de citações. Desenvolvimento de coleções.

Analysis of references in theses as support to decision making in the development of collections of the academical library

ABSTRACT

The work here presented accomplishes the study of the use of technical and scientific information, through the analysis of the references used in the elaboration of academic theses. The quantitative research uses bibliometrics to support the decision taking in the development of collections in a library. The group of extracted data of the protected theses is analyzed among the years from 1999 to 2006, guided by teachers belonging to the Department of Administrative Sciences of the School of Administration of the Federal University of Rio Grande do Sul (UFRGS) representing, until this date, a total of 80 theses available in complete text in Internet. The general objective is to create a database of the references bibliography of theses of a masters degree program and later to study the use of the technical and scientific information through the analysis of the citations. The specific objectives are to collect the available theses in digital format, that will serve as base for the analysis; to model and to develop a database of the bibliographical references used in all the theses; and to create queries in the database allowing to analyze the typology of the bibliographical materials used in the citations, its obsolescence, the titles of mentioned newspapers and the language of the consulted materials. The obtained results demonstrate that the information are useful for decision taking of the administrators' of the library, in what they concern the development of collections.

Key-Words: Database. Analysis of citations. Development of collections.

SUMÁRIO

LISTA DE FIGURAS	i
LISTA DE TABELAS	ii
1 INTRODUÇÃO	1
1.1 Objetivos.....	2
2 REFERENCIAL TEÓRICO	4
2.1 Bibliometria e o Desenvolvimento de Coleções.....	4
2.2 Banco de Dados	11
2.2.1 Modelo Entidade-Relacionamento	13
2.2.2 Modelo Relacional.....	16
3 METODOLOGIA.....	24
4 RESULTADOS E ANÁLISES	32
4.1 Análise da tipologia das referências	34
4.2 Análise da obsolescência das referências	35
4.3 Análise dos títulos de periódicos mais citados	37
4.4 Análise dos idiomas das referências	37
5 CONCLUSÕES.....	39
6 ANEXOS	42
7 REFERÊNCIAS BIBLIOGRÁFICAS	44

LISTA DE FIGURAS

Figura 2.1: Exemplo de diagrama entidade-relacionamento.....	14
Figura 2.2: Exemplo de modelo relacional.	18
Figura 2.3: Exemplo de comando de criação das tabelas Teses e Areas.	19
Figura 3.1: Modelo ER do banco de dados de teses.....	26
Figura 3.2: Mapeamento modelo ER-relacional feita no DBDesigner.	28
Figura 3.3: <i>Script</i> de criação do banco de dados de teses, na linguagem SQL.	29
Figura 4.1: Gráfico da relação de quantidade de teses por área.	32
Figura 4.2: Gráfico da relação de quantidade de orientações por professores.	33
Figura 4.3: Gráfico da relação de quantidade de teses publicadas por ano.....	33
Figura 4.4: Gráfico da distribuição das referências por tipos de publicação	35
Figura 4.5: Gráfico da distribuição de referências entre os anos de 1990 e 2006.....	36
Figura 4.6: Gráfico da distribuição das referências pelos idiomas.	38

LISTA DE TABELAS

Tabela 4.1: Distribuição das referências por tipos de publicação	34
Tabela 4.2: distribuição das obras por décadas	36

1 INTRODUÇÃO

No contexto universitário, as bibliotecas são consideradas como promotoras e divulgadoras da erudição e do conhecimento. Cabe à biblioteca universitária satisfazer as demandas informacionais de seus usuários para que eles desempenhem adequadamente suas atividades de ensino, pesquisa e extensão. Portanto, uma biblioteca com um acervo de qualidade reflete diretamente sobre os resultados esperados nas universidades.

Entretanto, os responsáveis por bibliotecas encontram dificuldades em melhorar de forma objetiva a tarefa complexa de planejamento. Uma das funções que mais necessita de atenção na administração de uma biblioteca é o desenvolvimento de coleções¹ (acervo), que, segundo Gutiérrez (1996), permite identificar as fortalezas e debilidades da coleção de uma biblioteca em termos das necessidades informacionais da sua comunidade.

Para que esta tarefa tenha um bom desempenho, algumas atividades específicas devem ser realizadas, tais como, alocação de recursos financeiros, seleção, aquisição, reposição e descarte de material informacional. Segundo Klaes (1991), para tanto, é importante coletar e organizar dados referentes ao comportamento do usuário para apoiar a tomada de decisão quanto ao desenvolvimento de coleções. As citações utilizadas na elaboração de teses oferecem dados que podem gerar informações valiosas para apoiar a tomada de decisão no desenvolvimento de coleções nas bibliotecas universitárias.

A quantidade de dados disponíveis nas citações de teses, que devem ser analisadas, geralmente é suficientemente grande para justificar o fato de ser necessário o desenvolvimento de ferramentas automáticas que auxiliem neste processo. Um banco de dados estruturado é parte fundamental desta ferramenta.

Sem uma ferramenta de apoio para organização dos dados, seria muito complexa a tarefa administrativa das bibliotecas universitárias de gerar informações realmente úteis para a tomada de decisão, no que diz respeito ao desenvolvimento de suas coleções.

No processo de desenvolvimento de coleções de uma biblioteca universitária, deve-se utilizar critérios quantitativos a fim de que se possa estabelecer um acervo capaz de atender às demandas informacionais de seus usuários.

¹ O acervo de uma biblioteca é formado por coleções de livros, coleções de periódicos etc, mas todo o acervo pode também ser chamado de coleção.

Vergueiro (1993) questiona se as coleções nas bibliotecas estão sendo realmente desenvolvidas com critérios e se elas estão seguindo qualquer tipo de parâmetro para seu desenvolvimento. A bibliometria pode dar suporte a estes critérios e parâmetros exigidos.

A bibliometria não é uma disciplina nova. Consiste em coletar um grande número de referências – citações² – para aplicar-lhes técnicas que facilitem seu entendimento, agrupando os pontos em comum nelas apresentados para submetê-las a uma análise. É uma ferramenta valiosa para entender os limites de um assunto que nos permite preparar estratégias (DOU, 2006). Dentre as técnicas bibliométricas está a análise de citações.

A análise das citações expressas nas dissertações e teses pode fornecer informações importantes no momento de selecionar quais são os títulos relevantes para compor as coleções de uma biblioteca universitária. A tipologia, a obsolescência e o idioma do material consultado, bem como os títulos de periódicos mais utilizados podem apoiar, por exemplo, a tomada de decisão quanto à distribuição dos recursos financeiros para investimento na aquisição de materiais informacionais e no armazenamento dos materiais obsoletos visando otimizar a utilização do espaço físico.

1.1 Objetivos

O objetivo geral desta pesquisa é criar um banco de dados das referências bibliográficas de teses de um programa de pós-graduação e posteriormente estudar o uso dessa informação técnica e científica por meio da análise das citações.

Espera-se que os resultados obtidos com esta pesquisa possam apoiar decisões administrativas no processo de desenvolvimento de coleções na biblioteca universitária.

Os objetivos específicos deste projeto de pesquisa são:

- a) coletar as teses disponíveis em formato digital, que irão servir de base para a análise;
- b) modelar e desenvolver um banco de dados das referências bibliográficas utilizadas em todas as teses;

²Existe distinção entre os termos referência e citação. Referências são todas as obras utilizadas como fontes em um trabalho. Citações são colocações no texto extraídas das obras referenciadas. Porém, desde que normatizado que todas as citações feitas num texto têm que ser referenciadas, os termos vêm sendo usados como sinônimos.

- c) criar consultas no banco de dados permitindo a análise de citações que identifique:
- i) a tipologia dos materiais bibliográficos utilizados nas citações; ii) a obsolescência dos materiais bibliográficos citados; iii) os títulos de periódicos citados; e iv) o idioma dos materiais consultados.

2 REFERENCIAL TEÓRICO

Neste capítulo pretende-se fazer uma descrição mais aprofundada sobre os assuntos que são fundamentais para o entendimento e desenvolvimento desta pesquisa. Na Seção 2.1 serão apresentadas as principais definições sobre bibliometria e o desenvolvimento de coleções, assim como as justificativas e importância destas para a tomada de decisão. Na Seção 2.2 são apresentadas as principais definições da área de banco de dados, tais como modelagem e implementação, que são aquelas utilizadas no desenvolvimento deste trabalho.

2.1 Bibliometria e o Desenvolvimento de Coleções

A bibliometria é uma área da ciência da informação, que tem como objetivo fazer um estudo sobre textos e/ou informações, e a análise de citações é um dos métodos usados para esse objetivo. A análise de citação estuda a produção bibliográfica de um ou vários autores, para que com isso possa criar métodos para classificar a importância de determinado trabalho, bem como definir critérios para comparação entre diferentes autores e trabalhos.

Não é novidade que, com o desenvolvimento da internet, transformações significativas vêm acontecendo em todos os setores da sociedade. As redes de informação trouxeram novas maneiras de trabalho, de diversão, de fazer negócios, de estudar, de comunicar, de pesquisar etc. Com a facilidade de comunicação e acesso a informação, principalmente de cunho científico, a bibliometria apresenta-se como um excelente mecanismo para auxiliar a análise dos dados, maximizando tanto a pesquisa de informações, quanto uma produção científica mais eficiente.

Para que determinado trabalho tenha validade e seja aceito pela comunidade científica, ele passa por um longo processo, desde seu desenvolvimento até sua publicação. A publicação é, para a maioria daqueles que atuam na pesquisa, não só um indicador de produtividade, mas o resultado final de uma cadeia de trabalho. Desta forma, para que uma contribuição tenha validade, ela necessita ser publicada, julgada e incorporada aos conhecimentos já existentes e reconhecidos previamente. Somente através da publicação é que o ciclo de conhecimento está completo, após a aceitação de outros cientistas dessa

nova descoberta. Por esse motivo é importante avaliar o conhecimento seguindo os passos da publicação científica de um ou vários pesquisadores de determinado ramo da ciência.

Os avanços tecnológicos presentes nos dias atuais facilitaram o acesso e tratamento da informação, entretanto a Bibliometria como ciência, não é uma ciência nova, o termo que a precedeu era *statistical bibliography*, utilizado pela primeira vez por E. Wyndham Hulme em 1922. Contudo, com o passar do tempo e com o amadurecimento da ciência surge a necessidade de utilizar um termo mais adequado. O termo *Bibliometrics* (Bibliometria) foi proposto por Pritchard em 1969, em seu trabalho “*Statistical bibliography, or bibliometrics?*”. Este é o marco considerado pelos principais autores para denominar a origem do termo. A definição que Pritchard utilizou foi a aplicação de métodos matemáticos e estatísticos a livros, artigos e outras mídias de comunicação.

Para Irvine e Martin (1989) indicadores científicos são cada vez mais necessários para os que formulam as políticas científicas no plano nacional, os quais devem determinar as prioridades de pesquisa entre e dentro dos mais diversos campos científicos. Os tomadores de decisões que pertencem tanto aos órgãos de governo quanto às agências financiadoras, precisam de dados sistemáticos sobre o desempenho das diversas áreas para poder escolher, com melhores fundamentos, onde concentrar os recursos financeiros e humanos limitados de que dispõem.

O conjunto de referências bibliográficas (citações) utilizadas na elaboração de um documento mostra o relacionamento de um documento com outro, evidenciando elos entre indivíduos, instituições e áreas de pesquisa. Ou seja, as citações que estão na bibliografia são importantes aparatos para que estudiosos possam amparar seus estudos e dão credibilidade para os fatos citados no texto, além de permitir aos pesquisadores da área a oportunidade de conhecer trabalhos que tratam do tema de seu interesse. Assim, a análise das citações de um trabalho contribui para avaliar a informação coletada pelo tipo de literatura utilizada, dirigir o leitor para outras fontes de informação sobre o assunto, além de contribuir para o reconhecimento de um cientista em particular.

A bibliometria, enquanto método quantitativo de investigação da ciência, utiliza a análise de citações como uma de suas ferramentas, a fim de medir o impacto e a visibilidade de determinados autores dentro de uma comunidade científica, verificando quais “escolas” do pensamento vigoram dentro das mesmas. Além disso, a análise de citações possibilita a mensuração das fontes de informação utilizadas, como o tipo de documento, o idioma e os periódicos mais citados. Utilizando estes indicadores, é possível

saber como se dá a comunicação científica de uma área do conhecimento, obtendo-se, assim, um “mapeamento” da mesma, descobrindo teorias e metodologias consolidadas (Vanz; Caregnato, 2003).

Existem duas classes em que os estudos de bibliografia são divididos: microescala e macroescala, essa divisão é baseada no tamanho do enfoque que o estudioso quer dar. Leta e Cruz (2003) exemplificam que os estudos em macroescala tem um enfoque nas atividades científicas de uma determinada área, uma nação ou região do mundo, enquanto que os estudos em microescala tem como base a análise de projetos, programas de pesquisas, instituições de ensino e pesquisa. Portanto, esse trabalho tem como escopo ser desenvolvido sobre a ótica e características de microescala.

No meio acadêmico, as bibliotecas tem um papel fundamental de prover e divulgar o conhecimento e a informação. É responsabilidade da biblioteca universitária suprir todas as necessidades informacionais de seus usuários, para que todos tenham base para desenvolver seus projetos de pesquisa e ensino.

A eficiência de uma instituição de ensino e pesquisa está diretamente ligada a qualidade de suas coleções, entretanto como Negrete Gutiérrez (1996) destaca que os responsáveis pelas bibliotecas encontram dificuldades em uma de suas principais funções que é o desenvolvimento de coleções, que engloba a identificação de pontos fortes e fracos da biblioteca no que se refere a necessidade de seus usuários. Apesar de que artigos, monografias, dissertações e teses representarem a produção científica, normalmente eles passam despercebidos para mapear o conhecimento.

Normalmente, os diretores de bibliotecas para orientar novas aquisições, solicitam aos membros dos departamentos da instituição que lhes enviem suas listas com os itens que gostariam de ver adquiridos, entretanto os próprios trabalhos científicos dos usuários já podem indicar os desejos de aquisição (OLIVEIRA, 2004). Portanto é altamente recomendável a utilização de ferramentas que proporcionem formas de maximizar a qualidade da biblioteca, e a análise de citações é uma das ferramentas que oferecem subsídios para que a coleção cresça de forma consistente, qualitativa e quantitativamente.

A análise de citações tem como um de seus objetivos desenvolver e controlar o conhecimento dentro de uma biblioteca universitária. Tarapanoff (2001) afirma que ela é um dos pilares da análise de ambiente, contribuindo principalmente em oferecer informações confiáveis para oferecer suporte no processo decisório, sobretudo porque não é somente no momento final que o processo decisório se resume, ele é longo e passa

por várias etapas, como reflexão dos problemas, investigação e compilação dos dados e finalmente a análise desses dados.

A tomada de decisão é um dos três processos de uso estratégico da informação. Durante a tomada de decisões, a principal atividade é o processamento e a análise da informação a partir das alternativas disponíveis, cujas vantagens e desvantagens são pesadas (CHOO, 2003). Sendo assim, o estudo de citações utilizadas na elaboração das teses pode ajudar, por exemplo, a tomada de decisão para determinar a alocação mais apropriada dos recursos financeiros disponíveis para investimento entre os periódicos e as monografias que farão parte da coleção-núcleo da biblioteca.

O desenvolvimento de coleções – processo que permite às bibliotecas construir coleções de materiais documentais que correspondam às necessidades de informação de seus usuários – começou a difundir-se há mais de 30 anos nas bibliotecas universitárias de países anglo-saxões, em grande parte em virtude de restrições orçamentárias. Mesmo sem restrições orçamentárias, o crescimento exponencial das coleções e a carência de espaço físico, obrigavam os bibliotecários a implementar políticas mais precisas de descarte e armazenamento (NEGRETE GUTIÉRREZ, 1993).

Por outro lado, com as restrições orçamentárias e o encarecimento das publicações, os responsáveis pelo desenvolvimento de coleções precisavam justificar a necessidade de aquisição e a racionalizar os recursos disponíveis para investimento.

O desenvolvimento de coleções é importante porque permite aos administradores das bibliotecas antecipar-se às necessidades e demandas informacionais da comunidade usuária através de planejamento, que é uma tarefa de fundamental importância em toda as áreas administrativas; e adequar-se aos ambientes internos e externos em que a biblioteca está inserida, tais como: direção da unidade, política, legislação, mudanças sociais ou demográficas, mercado editorial, inovações tecnológicas e padrões de desempenho exigidos (DIAS; PIRES, 2003). Choo (2003) sustenta a afirmação anterior quando conclui que a organização usa a informação para dar sentido às mudanças do ambiente externo.

O desenvolvimento de coleções no contexto da biblioteca universitária requer a compreensão das necessidades informacionais dos autores das teses. Estas demandas podem ser difíceis de discernir, mas a análise de citações possibilitará a compreensão clara da natureza das necessidades informacionais dos pesquisadores (WILLIAMS; FLETCHER, 2006).

Entre os métodos e as técnicas para tomada de decisão, Dias e Pires (2003)

apresentam a bibliometria como uma ferramenta que elabora indicadores de tendências, gráficos, figuras e mapas que sintetizarão as informações para a tomada de decisão.

Com o objetivo de dar apoio a tomada de decisões dos administradores de bibliotecas para o adequado planejamento do desenvolvimento de coleções da biblioteca universitária, este trabalho propõe-se a investigação do uso da informação técnica e científica por meio da análise das citações expressas na elaboração das teses acadêmicas, como um mecanismo que permita adequar a coleção da biblioteca às necessidades informacionais de seus usuários. Esta pesquisa vale-se do estudo bibliométrico para apoiar a tomada de decisão no desenvolvimento de coleções.

A bibliometria estuda as características quantitativas da produção, disseminação e uso da informação registrada através do desenvolvimento de padrões e modelos matemáticos que permitem mensurar esses processos. Possibilita situar desde a produção intelectual de um país em relação ao mundo até a produção dos cientistas em relação à sua própria comunidade. Pode ter por objeto de estudo documentos, autores e usuários; como variáveis, as citações e por objetivos, a alocação de recurso, tempo e dinheiro (MACIAS-CHAPULA, 1998).

Diversos autores discutem a utilidade e limitações da análise de citações: Garfield (1977) afirma que a frequência da citação pode determinar a seleção da coleção-núcleo dos periódicos, mas considera este critério insuficiente para determinar todas as seleções. Le Coadic (2004) nos esclarece que ao citar o autor pode ter outras motivações, tais como: citação-recompensa, citação política, citação-álibi, citação-persuasão e a autocitação. Meadows (1999) destaca que a validade de examinar as citações como medida de qualidade requer uma classificação destas em pelo menos três grupos: essenciais, suplementares e negativas. Smith (1981) nos alerta que a citação não é responsável pela qualidade ou importância dos materiais citados. Spinak (1998) concorda com Smith quando afirma que os números refletem a quantidade ou atividade e de certa maneira a produtividade, mas não a qualidade e conteúdo das publicações.

Porém, o uso da análise de citações pode, por exemplo, apresentar fórmulas de alocação orçamentária para determinar a distribuição do orçamento disponível, na suposição que a proporção dos recursos alocados entre a coleção de monografias e a coleção de periódicos deve igualar-se às proporções das citações destes materiais feitas pelos pesquisadores em seus trabalhos (DEVIN; KELLOGG, 1990).

Macias-Chapula (1998, p. 136) destaca as 15 funções específicas para citar, todas

positivas, de acordo com Weinstock³:

(1) prestar homenagem aos pioneiros; (2) dar créditos para trabalhos relacionados; (3) identificar metodologia, equipamento, etc.; (4) oferecer leitura básica; (5) retificar o próprio trabalho; (6) retificar o trabalho de outros; (7) analisar trabalhos anteriores; (8) sustentar declarações; (9) informar aos pesquisadores de trabalhos futuros; (10) dar destaque a trabalhos pouco disseminados, inadequadamente indexados ou desconhecidos (não citados); (11) validar dados e categorias de constantes físicas e de fatos, etc.; (12) identificar publicações originais [...]; (13) identificar publicações originais que descrevem conceitos ou termos epônimos⁴ [...] (14) contestar trabalhos ou idéias de outros; e (15) debater a primazia das declarações de outros.

Segundo Line (1978), as listas de periódicos oriundas dos estudos de citações não devem representar o fator preponderante no momento da compra, mas podem ser úteis para identificar os periódicos dignos de uma avaliação mais acurada pelos bibliotecários familiarizados com as necessidades de seus usuários.

No entanto, em uma situação ideal, o desenvolvimento de coleções deve ser executado por um bibliotecário que seja também especialista no assunto, na ausência de um, a análise de citações pode fornecer uma orientação útil para o bibliotecário não especializado (BROADUS, 1977).

Mas de acordo com alguns autores, apesar das limitações, o estudo analítico das citações representa uma importante ferramenta de apoio a decisões administrativas, quais sejam: (a) armazenar a coleção em depósitos por sua antiguidade (ACKERSON, 2001); (b) distribuir os recursos financeiros disponíveis para investimento entre a coleção de periódicos e de monografias (DEVIN; KELLOG, 1990); (c) selecionar quais os idiomas das obras que farão parte do acervo (KNIEVEL; KELLSEY, 2005); (d) avaliar o uso de uma coleção de periódicos (SYLVIA, 1998); (e) decidir sobre a gerência da coleção (GREENE, 1994); (f) analisar atividades de pesquisa realizadas na instituição (SMITH, 1981); (g) estudar os usuários como um método indireto de análise (URBANO SALIDO, 2001); e (h) escolher o tipo de material – por exemplo, literatura cinzenta – que deve fazer parte da coleção da biblioteca (URBANO SALIDO, 2000).

Embora nem sempre seja plenamente perceptível, o uso da

³Macias-Chapula (1998) não fornece a referência de Weinstock. Meadows (1999, p. 90) trata do assunto e apresenta a seguinte referência: WEINSTOCK, M. Citation Index. **Encyclopaedia of Library and Information Science**, 5, p. 16-40 (1971).

⁴Que dá ou empresta seu nome a alguma coisa.

literatura cinzenta tem sido uma constante nas ciências sociais aplicadas, especialmente em administração de empresas e ciências contábeis. As informações disponíveis sob a forma de relatórios, estudos setoriais e evolução de índices diversos disponíveis em sites oficiais ou organizações não governamentais têm sido utilizadas para fundamentar análises e trabalhos acadêmicos publicados em revistas científicas, anais de congressos ou livros. Da mesma forma, citações de teses e dissertações são uma constante, embora essas publicações também sejam consideradas como parte da literatura cinzenta (CÓRTEZ, 2006).

A biblioteca universitária, um dos ambientes do grande sistema de informação, participa do processo de comunicação científica. E como explica Spinak (1998), a bibliometria pode colaborar analisando o aspecto mais relevante deste sistema: a comunicação.

[...] Entre os sistemas mais importantes em que está imerso o cientista estão: o ambiente cultural, o ambiente político, os colégios invisíveis, as organizações formais, os grupos de trabalho, o sistema legal e econômico, as associações profissionais e finalmente o sistema de informação. A bibliometria é, pois, uma disciplina com alcance multidisciplinar e que analisa um dos aspectos mais relevantes e objetivos desta comunidade [científica], a comunicação impressa (SPINAK, 1998, p. 142, tradução nossa, grifo nosso).

Como ferramenta auxiliar do planejamento, a bibliometria pesquisa o uso dos documentos, buscando padrões quantitativos para análise. Tal pesquisa pode ser feita em relação às citações.

A bibliometria oferece, a partir do estabelecimento de padrões quantitativos, modelos de comportamento, não só do conhecimento ou da literatura, mas também das necessidades informacionais centradas no usuário, acumulando dados que, devidamente agrupados, auxiliam o processo de tomada de decisão (TARAPANOFF; MIRANDA; ARAÚJO JÚNIOR, 1995, p. 97).

Métodos quantitativos são instrumentos que devem ser considerados pelos bibliotecários no processo de tomada de decisão. Nesse sentido, Figueiredo e Lima destacam que, no desenvolvimento de coleções, a interpretação das análises quantitativas de coleções são instrumentos auxiliares na tomada de decisão. Relacionam-se com planejamento, seleção, revisão e desbastamento, sendo impraticável desenvolver uma coleção que reúna todos os itens publicados em qualquer área do conhecimento

(FIGUEIREDO; LIMA, 1998).

Muitas bibliotecas acadêmicas começaram a observar as citações apresentadas nas teses produzidas por seus alunos como evidência das habilidades desenvolvidas por eles para pesquisar. Elas servem também como um indicador parcial da qualidade das obras utilizadas por eles na elaboração de suas monografias. Estudá-las, tem ajudado as bibliotecas universitárias a conduzir o desenvolvimento de suas coleções. Citações são particularmente atraentes porque suas informações podem ser descritas e contadas e seus dados podem ser coletados com relativa facilidade, sendo que esta coleta não é contaminada pela participação de respostas e opiniões (TUÑÓN; BRYDGES, 2005).

Isto posto, conclui-se que a análise das citações utilizadas na elaboração de teses acadêmicas constitui-se em uma ferramenta importante para apoiar a tomada de decisão quanto ao desenvolvimento de coleções em bibliotecas universitárias.

Um problema, porém, surge devido a grande quantidade de dados que são necessários para realizar a análise de citações de teses. A tarefa de fazer esta análise sem o auxílio de alguma ferramenta computacional eficiente tornar-se-ia exaustiva. Por isso, justifica-se a necessidade de desenvolvimento de ferramentas automáticas que auxiliem este processo. Um banco de dados estruturado é parte fundamental desta ferramenta.

2.2 Banco de Dados

Banco de dados é uma coleção de dados que têm alguma relação, que são agrupados em uma estrutura regular para armazenar informações. Normalmente essas informações são relacionadas entre si, e tem como objetivo um mesmo tópico ou propósito específico. Os dados são fatos que podem ser gravados e que tenham algum significado implícito.

Elmasri e Navathe (2005) exemplificam que os dados podem ser nomes, números telefônicos e endereços de pessoas que você conhece, e essas informações podem ter sido escritas em uma agenda de telefones ou armazenadas de alguma forma em um computador, essas informações são uma coleção de dados com um significado implícidos, conseqüentemente, um banco de dados.

A definição apresentada anteriormente é muito genérica, entretanto o uso do termo banco de dados é mais restrito, Elmasri e Navathe (2005) listam as seguinte propriedades implícitas para se determinar um banco de dados:

- Um banco de dados representa alguns aspectos do mundo real, sendo chamado, às vezes de mini-mundo. As mudanças no mini-mundo são refletidas no banco de dados.
- Um banco de dados é uma coleção lógica e coerente de dados com algum significado inerente. Uma organização de dados ao acaso (randômica) não pode ser corretamente interpretada como um banco de dados.
- Um banco de dados é projetado, construído e povoado por dados, atendendo a uma proposta específica. Possui um grupo de usuários definido e algumas aplicações preconcebidas, de acordo com o interesse desse grupo de usuários.

Em outras palavras, um banco de dados deve possuir algumas fontes das quais os dados são derivados, deve possuir em maior ou menor nível de interação com o mundo real e um público interessado em seu conteúdo.

A evolução tecnológica, os meios eletrônicos cada vez mais difundidos e a necessidade de se ter controle de todos os dados, seja de uma pessoa física ou de uma empresa multinacional, tem gerado um aumento significativo no volume de dados trabalhados, o desenvolvimento de um banco de dados é especialmente recomendável nesses casos, quando o alto volume de dados impossibilita o aproveitamento real da informação. Goebel e Gruenwald (1999) sintetizam que todo esse conjunto de dados contém uma preciosa quantidade de informações, entretanto sem ferramentas que auxiliem o armazenando e a análise esses dados são desperdiçados, não gerando nenhuma informação real (útil).

Os dados por si só não geram conhecimento, eles, se olhados isoladamente, não têm significado relevante, um dado é dentre outros uma representação de algum evento do mundo físico, de um fato, um objeto ou uma propriedade, somente depois de identificar, armazenar, organizar, agrupar e recuperar os dados é possível ser realizada uma interpretação, gerando assim uma informação útil. Em outras palavras, para ser considerada uma informação, o dado tem que ser contextualizado. Num banco de dados esta contextualização é alcançada com a estruturação correta dos dados. A partir da interpretação e utilização das informações pode-se obter algum tipo de conhecimento.

O banco de dados é um recurso valioso para a maioria das organizações, ele pode guardar um grande volume de informações e proporcionar um ambiente para recuperação e armazenamento de dados (SILBERSCHATZ *et al.*, 1999).

O processo de criação de um banco de dados tem início na tarefa de modelagem dos dados. Para tanto, um trabalho anterior de análise do sistema deve ter sido realizado junto ao usuário final do sistema. Resumidamente, o usuário, juntamente com o analista de sistemas, definem quais dados são importantes para solucionar o problema, neste caso, é gerada uma lista de requisitos dos dados do sistema. Com base nestes requisitos, o projetista de banco de dados deve modelar o banco de dados.

A modelagem do banco de dados normalmente é feita em duas etapas, sendo a primeira a criação de um modelo de alto nível de abstração, como por exemplo o modelo Entidade-Relacionamento; e a segunda etapa uma modelagem física, por exemplo usando o modelo Relacional, aonde o modelo em alto nível é transformado em um modelo para a real implementação em um Sistema Gerenciador de Banco de Dados (SGBD).

2.2.1 Modelo Entidade-Relacionamento

O modelo Entidade-Relacionamento (ER) foi proposto em 1976 por Peter Chen, como uma ferramenta adicional no projeto de banco de dados, é uma fase importante no planejamento de uma aplicação de um banco de dados bem sucedido (SILBERSCHATZ *et al.*, 1999).

O modelo caracteriza-se por ter apresentar um alto grau de independência e detalhes da implementação bem como grande afinidade com o mundo real, e sua representação se baseia em um protocolo gráfico chamado Diagrama de Entidade-Relacionamento, a ligação com o mundo real se dá por um conjunto de objetos chamados entidades e pelo conjunto dos relacionamentos entre esses objetos..

A Figura 2.1 exemplifica um trecho do banco de dados das referências de teses, com as entidades TESES e REFERENCIAS, e o relacionamento POSSUI que representa uma associação entre as duas. Neste exemplo pode-se observar algumas características (atributos) das entidades.

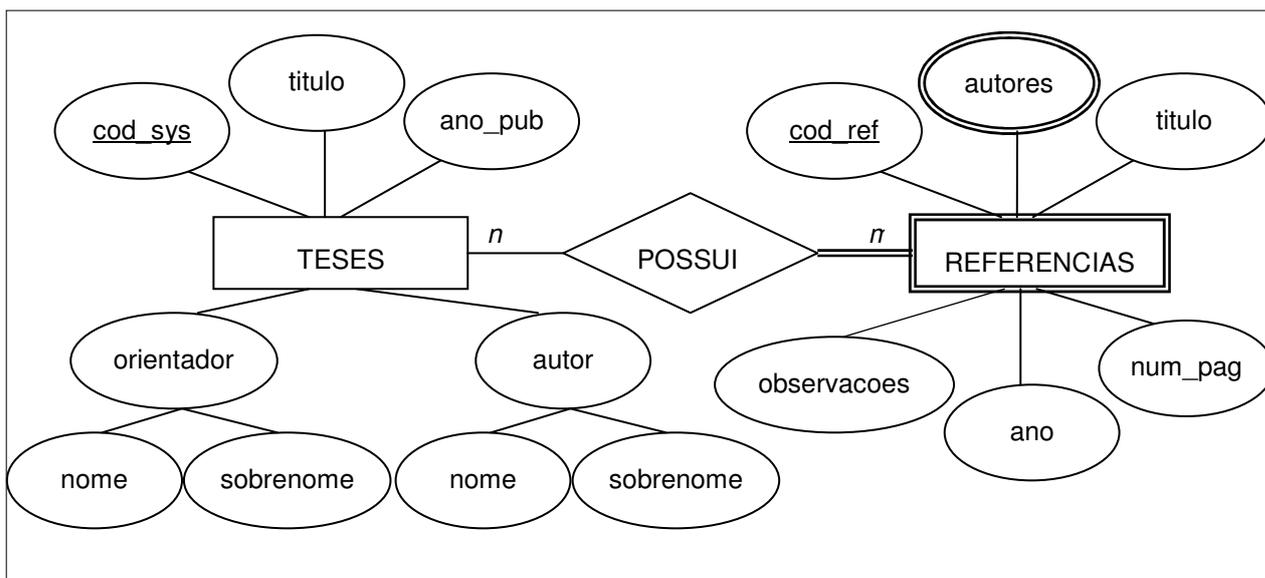


Figura 2.1: Exemplo de diagrama entidade-relacionamento.

Fonte: elaborada pelo autor.

Date (2005) define entidade como uma coisa que pode ser identificada distintamente, e continua classificando em entidades regulares (ou simplesmente entidades) e entidades fracas. Uma entidade fraca é uma entidade que depende de outra pra existir, no sentido de que ela não pode existir se a outra não existir, enquanto que entidade regular pode existir por si só. Por exemplo, na Figura 2.1 a entidade REFERENCIAS está ligada a entidade TESES, e é uma entidade fraca, porque não existe uma referência sem que haja uma tese, no que se refere ao banco de dados. Neste caso, se determinada tese for excluída as referências que aparecem nela também devem ser excluídas, ao contrário, TESES é uma entidade regular e existe por si só, independente de todas as outras.

Como dito anteriormente, o modelo baseia-se em um protocolo gráfico para tratar a modelagem em alto nível. Dentre as opções de gráficos descritas por Silberschatz *et al.* (1999), as mais importantes são:

- Retângulos: Representam entidades. Na Figura 2.1 TESES e REFERENCIAS são entidades.
- Elipses: Representam atributos.
- Elipses duplas: Representam atributos multivalorados.
- Losangos: Representam relacionamentos entre entidades. Na Figura 2.1 POSSUI é um relacionamento entre TESES e REFERENCIAS.
- Linhas Simples: Representam a união de entidades a atributos e de entidades aos seus relacionamentos.

- Linhas Duplas: Representam participação total (Dependência de Existência).
- Atributos Sublinhados: Representam chaves. Na Figura 2.1 “cod_sys” é a chave da entidade TESES e “cod_ref” é chave da entidade REFERENCIAS.

Uma entidade representa um objeto real como por exemplo uma pessoa, ou um objeto abstrato como por exemplo um empréstimo, porém ela deve ter uma relevância significativa para o negócio. As entidades possuem propriedades, chamados de atributos, que descrevem características da entidade. É interessante que os atributos também representem informações referentes ao negócio. Eles podem ser classificados da seguinte forma: (DATE, 2005).

Simples – Atributo não é divisível. No exemplo da Figura 2.1, “titulo” e “ano_pub” são atributos simples da entidade TESES.

- **Composto** – Atributo pode ser dividido em atributos simples. No exemplo da Figura 2.1, “orientador” e “autor” são atributos compostos da entidade TESES.
- **Monovalorado** – Atributo tem valor único para uma determinada entidade. No exemplo da Figura 2.1, todos exceto “autores” são atributos monovalorados da entidade REFERENCIAS.
- **Multivalorado** – Atributo tem um limite inferior e superior de valores permitidos para cada entidade. No exemplo da Figura 2.1, “autores” é um atributo multivalorado da entidade REFERENCIAS.

Os relacionamentos possuem restrições que limitam as possibilidades de combinações de entidades que podem participar do conjunto de relacionamentos correspondentes. Para descrever essa restrição utiliza-se a cardinalidade, que pode ser (ELMASRI; NAVATHE, 2005):

- **Um para um (1:1)** – Uma entidade em A está associada no máximo a uma entidade em B, e uma entidade em B está associada a no máximo uma entidade em A;
- **Um para muitos (1:n)** – Uma entidade em A está associada a várias entidades, em B, e uma entidade em B, entretanto, deve estar associada no máximo a uma entidade em A;
- **Muitos para um (n:1)** – Uma entidade em A está associada a no máximo uma entidade em B, e uma entidade em B pode estar associada a várias entidades em A;
- **Muitos para muitos (n:m)** – Uma entidade em A está associada a várias entidades em B e uma entidade em B está associada a várias entidades em A.

2.2.2 Modelo Relacional

O modelo relacional surgiu devido a seguintes necessidades da época: aumentar a independência de dados nos sistemas gerenciadores de banco de dados; prover um conjunto de funções apoiadas em álgebra relacional para armazenamento e recuperação de dados; permitir processamento *ad hoc*. Este modelo foi resultado de um estudo teórico realizado por Codd (1970), tendo por base a teoria dos conjuntos e álgebra relacional. O modelo foi apresentado num artigo publicado em 1970, mas que só nos anos 80, foi implementado.

O modelo relacional se baseia na teoria dos conjuntos e na lógica de predicados de primeira ordem. O modelo representa o banco de dados como uma coleção de relações, aonde cada relação se parece com uma tabela de valores (ELMASRI; NAVATHE, 2005).

De acordo com Neves (2002) e Silberschatz *et al.* (2005), o modelo relacional é um dos modelos mais difundidos e é o mais utilizado para aplicações comerciais de processamento de dados em todo o mundo. Este modelo está sendo indicado para resolver a maioria dos problemas, e conquistou sua posição de destaque devido à sua simplicidade e flexibilidade que facilita o trabalho do programador, comparado com os modelos de dados anteriores a ele.

Um banco de dados relacional consiste em uma coleção de tabelas, cada uma com um nome único atribuído. Uma linha ou tupla numa tabela representa uma relação entre um conjunto de valores. Informalmente, uma tabela é um conjunto de entidades, e uma linha é uma entidade. Como uma tabela é uma coleção dessas relações, existe uma íntima correspondência entre o conceito de tabela e o conceito matemático de relação, do qual o modelo de dados relacional extrai seu nome (SILBERSCHATZ *et al.*, 2005).

Ainda segundo Silberschatz *et al.* (2005), os matemáticos definem uma relação com sendo um subconjunto de um produto cartesiano de uma lista de domínios, sendo essa definição quase a mesma definição de tabela, tendo como diferença que é matemáticos atribuem números para representar a lista de domínio, enquanto que no modelo relacional atribuímos nomes a atributos.

Pode-se descrever o banco de dados relacional de uma maneira informal ou mais formal. Date (2005) destaca que na descrição informal deve-se estar preocupado com aspectos práticos e usa-se os termos: tabela, linha e coluna, enquanto que na descrição

formal deve-se estar preocupado com a semântica formal, e usa-se os termos: relação (tabela), tupla (linha) e atributo (coluna).

O modelo relacional é definido por Date (2005) como tendo três aspectos, que são:

- Aspecto estrutural: os dados no banco de dados são percebidos pelo usuário como tabelas, e nada além de tabelas.
- Aspecto de integridade: essas tabelas satisfazem a certas restrições de integridade.
- Aspecto manipulador: os operadores disponíveis para que o usuário possa manipular essas tabelas, sendo os operadores de restrição, projeção e junção os mais importantes.

Em suma, o modelo de dados relacional consiste basicamente em um conjunto de estruturas de dados (tabela), uma linguagem de manipulação de banco de dados e regras de integridade que são definidas no esquema do banco de dados, e são aplicadas nas instâncias do banco de dados.

A tabela é o local aonde os dados são armazenados no banco de dados relacional. Uma tabela é um conjunto estruturado de linhas e colunas. Na tabela, cada linha contém um mesmo conjunto de colunas, no banco de dados podem existir de 1 a N tabelas, sendo limitada pelo *software* e do *hardware* utilizado. As tabelas mantêm associação entre si utilizando regras de relacionamento, aonde se associa um atributo de uma tabela a outra tabela.

No modelo relacional existe uma série de restrições que são implementadas e definem algumas regras de funcionamento dos sistemas gerenciadores de bancos de dados baseados neste modelo. Algumas destas restrições são baseadas no conceito de integridade. Segundo Date (2005), o termo integridade refere-se à precisão ou correção de dados no banco de dados. Nesse contexto “integridade” significa semântica e são as restrições de integridade que representam o significado dos dados.

A restrição de integridade da entidade é especificada em relações individuais e declara que nenhum valor da chave primária pode ser nulo. Isso se justifica porque o valor da chave primária implica que não podemos identificar algumas tuplas, Elmasri e Navathe (2005). Por exemplo, no caso de duas ou mais tuplas tivessem nulos para suas chaves primárias, não seria possível distinguir entre elas.

A restrição de integridade referencial é especificada entre duas relações e é utilizada para manter a consistência entre tuplas de duas relações. Informalmente, a

restrição de integridade referencial declara que uma tupla em uma relação que se refere a uma outra relação deve se referir a uma tupla existente naquela relação (ELMASRI; NAVATHE, 2005).

Para estabelecer as restrições de integridade e estabelecer os relacionamentos entre as tabelas, utiliza-se um campo identificado como chave. Uma chave primária é uma ou várias colunas que não possuem valores duplicados dentro de uma tabela. Uma chave estrangeira é uma ou várias colunas em que os valores estejam identificados necessariamente como chave primária de outra tabela. A chave estrangeira é o mecanismo que define os relacionamentos em um banco de dados relacional.

A Figura 2.2 apresenta um exemplo de tabelas do modelo relacional.

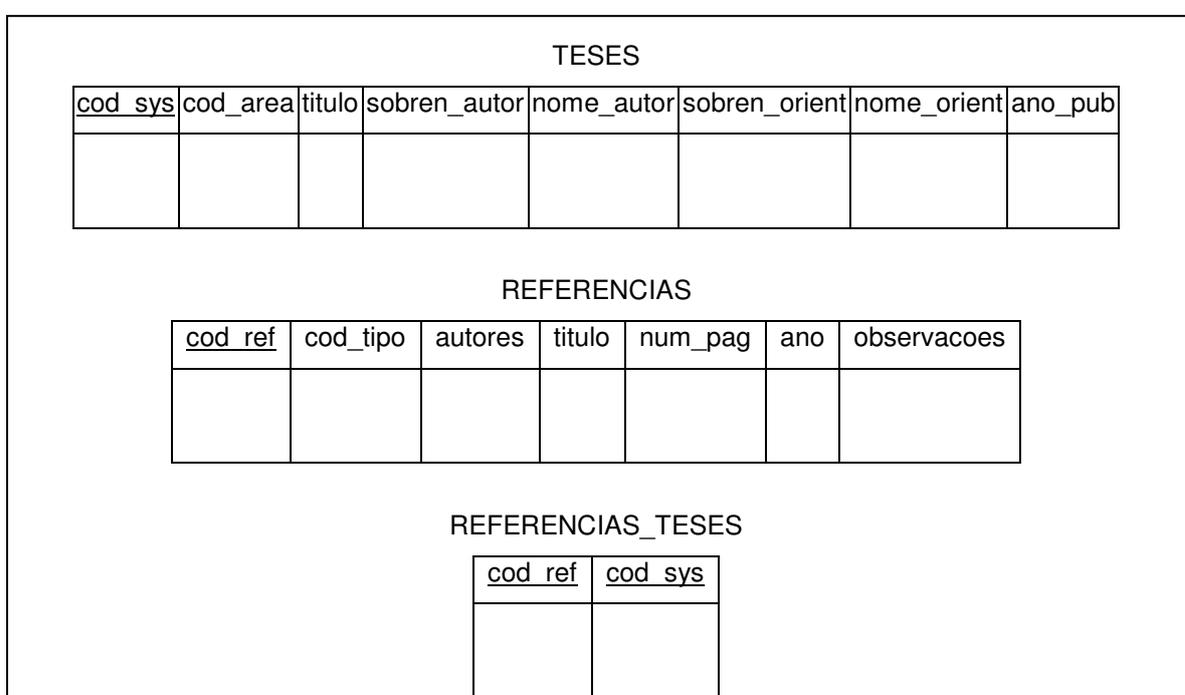


Figura 2.2: Exemplo de modelo relacional.
Fonte: elaborada pelo autor.

No exemplo ilustrado na Figura 2.2 podem ser observadas três tabelas: TESES, REFERENCIAS e REFERENCIAS_TESES. Na tabela TESES o atributo “cod_sys” é a sua chave primária. Na tabela REFERENCIAS o atributo “cod_ref” é a chave primária. Na tabela REFERENCIAS_TESES há um exemplo de chave primária composta formada pelos atributos “cod_sys” e “cod_ref”. Além disso, estes dois últimos atributos são chaves estrangeiras das demais tabelas.

Uma vez definidas no projeto físico quais são as tabelas que deverão ser criadas, a

próxima etapa do processo de criação de um banco de dados é utilizar um Sistema Gerenciador de Banco de Dados (SGBD), criar o banco de dados e criar as tabelas. A Figura 2.3 ilustra como poderia ser um comando de criação de uma tabela num banco de dados.

```

CREATE TABLE TESES (
    COD_SYS                CHAR(6)          NOT NULL,
    COD_AREA               CHAR(6)          NOT NULL,
    TITULO_TESES           VARCHAR(300)     NOT NULL,
    SOBRENOME_AUTOR        VARCHAR(50)     NOT NULL,
    NOME_AUTOR             VARCHAR(100)    NOT NULL,
    SOBRENOME_ORIENT       VARCHAR(50)     NOT NULL,
    NOME_ORIENT            VARCHAR(100)    NOT NULL,
    ANO_PUBLICACAO         CHAR(4)         NOT NULL,
    PRIMARY KEY (COD_SYS),
    FOREIGN KEY (COD_AREA) REFERENCES AREAS (COD_AREA));

CREATE TABLE AREAS (
    COD_AREA               CHAR(4)          NOT NULL,
    DESCRICAO_AREA         VARCHAR(100)     NOT NULL,
    PRIMARY KEY (COD_AREA));

```

Figura 2.3: Exemplo de comando de criação das tabelas Teses e Areas.

Fonte: Elaborada pelo autor.

Na Figura 2.3 pode-se observar o uso de alguns conceitos do modelo relacional e da linguagem SQL⁵. O comando “CREATE TABLE TESES” cria uma tabela chamada TESES que irá armazenar informações sobre todas as teses coletadas, e possui várias características que a representam, essas características são chamados atributos. São exemplos de atributos da tabela TESES: COD_SYS, COD_AREA, TITULO_TESES, SOBRENOME_AUTOR, NOME_AUTOR, SOBRENOME_ORIENT, NOME_ORIENT, ANO_PUBLICACAO. Após a inserção de dados nesta tabela, cada linha da tabela conteria um conjunto de todos esses dados, ou seja, todos os dados referentes a alguma tese estaria na mesma linha, sendo cada atributo em uma coluna.

Ainda utilizando o exemplo apresentado na Figura 2.3, a tabela TESES contém um campo especial chamado COD_SYS que é a representação de sua chave primária, esse campo tem o propósito de manter a integridade do sistema adicionando um índice, para

⁵A linguagem SQL (*Structured Query Language*) é a linguagem de consultas padrão utilizada no modelo relacional.

que cada registro seja único e possa ser recuperado. A tabela TESES interage com outras tabelas, dentre elas, a tabela AREAS que representa áreas de pesquisa e possui os atributos: COD_AREA e DESCRICAO_AREA, sendo COD_AREA sua chave primária. Toda tese foi desenvolvida em alguma área, ou seja, existe um relacionamento entre as tabelas TESES e AREAS representado pelo campo em comum entre as duas tabelas, que é uma chave estrangeira. Na tabela TESES o atributo COD_AREA é uma chave estrangeira que faz referência à tabela AREAS.

A utilidade do banco de dados existe em função de todo um conhecimento e tecnologia que foi sendo aperfeiçoada ao longo dos anos e todo esse desenvolvimento é encarnando em um tipo de *software* especializado chamado Sistema Gerenciador de Banco de Dados ou SGBD ou, ainda em termos mais coloquias sistema de banco de dados. Garcia-Molina *et al.* (2002) caracteriza o SGBD por ser uma ferramenta poderosa para criar e gerenciar grandes quantidades de dados de forma eficiente e permitir que esses dados persistam durante longos espaços de tempo com segurança, e complementa que esses sistemas estão entre os tipos mais complexos de *software* disponíveis.

Um SGBD é uma coleção de programas que permitem ao usuário definir, construir e manipular bancos de dados para as mais diversas finalidades, sendo o responsável pela persistência, organização e recuperação desses dados. Silberschatz *et al.* (1999) define SGBD como sendo uma coleção de dados inter-relacionados e um conjunto de programas para acessar esses dados, sendo seu principal objetivo fornecer uma maneira de recuperar informações do banco de dados que seja tanto conveniente quanto eficiente. O autor continua descrevendo como sendo sistemas projetados para gerenciar grandes blocos de informação, sendo que esse gerenciamento envolve definir estruturas para o armazenamento de informação e fornecer mecanismos para a manipulação de informação.

Nos dias atuais existem bancos de dados em praticamente todos os segmentos da sociedade e interage-se diariamente com eles, Silberschatz *et al.* (1999) destaca que embora as interfaces de usuário ocultem os detalhes do acesso a um banco de dados e a maioria das pessoas nem mesmo tenha consciência de estar lidando com um banco de dados, acessar bancos de dados é uma parte da vida de quase todo mundo.

Um SGBD deve possuir certas funcionalidades e capacidades que permitam ao seus usuários atingir os objetivos relacionados ao projeto, administração e uso de um grande banco de dados multiusuários. As funcionalidades e capacidades essenciais apresentadas por Garcia-Molina *et al.* (2002), Elmasri e Navathe (2005), que os SGBD devem possuir

são:

i) Controle de redundância e inconsistência de dados: O SGBD deve ser capaz de garantir que os dados não tenham duplicidade. Utilizando a chave primária não seria possível incluir dois registros com o mesmo código e nem excluir registros que tivesse um relacionamento com outras tabelas (chave estrangeira).

ii) Controle de acesso: O SGBD deve possuir restrição eficiente para controlar acesso não autorizado, aonde nem todos os usuários do sistema de banco de dados possam fazer tudo, e isso é conseguido através de privilégios, em que dependendo do privilégio recebido o usuário pode ter maior ou menor acesso às tabelas, bem como da permissão de quais operações pode realizar.

iii) Processamento eficiente de consultas: Os SGBD devem possuir estruturas de armazenamento eficientes que possibilitem fornecer funcionalidades para a execução de atualizações e consultas.

iv) Restrição de integridade: A maioria das aplicações de banco de dados possuem certas restrições de integridade que devem complementar os dados. O SGBD deve prover funcionalidades para a definição e garantia dessas restrições.

v) Armazenamento persistente: O SGBD deve possibilitar o armazenamento de quantidades muito grandes de dados que devem existir independentemente de quaisquer processos que estejam utilizando esses dados.

vi) Gerenciamento de transações: Um SGBD admite o acesso concorrente a dados, isto é, o acesso simultâneo por muitos processos distintos (chamados “transações”) ao mesmo tempo. Para evitar consequências indesejáveis do acesso simultâneo, o SGBD deve possuir o isolamento, a aparência de que as transações.

vii) Atomicidade: O SGBD deve fornecer mecanismo em que as transações sejam executadas completamente ou não sejam executadas de forma alguma.

viii) Resiliência: O SGBD deve possuir a capacidade de se recuperar de muitos tipos de falhas.

ix) Interface de programação: Um SGBD deve permitir ao usuário acessar e modificar dados através de uma linguagem de consulta poderosa.

O usuário deve possuir formas de acessar e modificar dados, isso se dá através de uma linguagem padrão específica, que é a *Structured Query Language*, ou Linguagem de Consulta Estruturada ou ainda simplesmente SQL.

A IBM desenvolveu a versão original da SQL, originalmente chamada Sequel,

como parte do projeto R do início da década de 1970, desde então a linguagem Sequel evoluiu, e seu nome mudou para SQL, e hoje ela se estabeleceu claramente como a linguagem padrão de banco de dados relacional (DATE, 2005).

Frequentemente a SQL é conhecida como sendo uma linguagem de consulta, mas conforme afirma Silberschatz *et al.* (1999), ela pode fazer muito mais do que apenas consultar um banco de dados, ela pode definir a estrutura dos dados, modificar dados no banco de dados e especificar restrições de segurança.

Os comandos da SQL são especificados por um padrão publicado pelo *American National Standards Institutes* (ANSI) e a *Internacional organization for Standardization* (ISO). A linguagem é dividida em três partes, uma linguagem de definição de dados (DDL – *Data Definition Language*), uma linguagem de manipulação de dados (DML – *Data Manipulation Language*) e a linguagem de controle de dados (DCL – *Data Control Language*). Silberschatz *et al.* (1999) descreve a DDL como sendo a linguagem que fornece comandos para definir esquemas de relação, excluir relações e modificar esquemas; e descreve a DML como a que fornece comandos para consultar, inserir e modificar tuplas no banco de dados.

O PostgreSQL⁶ é um SGBD Objeto-Relacional desenvolvido no Departamento de Ciência da Computação da Universidade da Califórnia em Berkeley. O PostgreSQL descende do código original do POSTGRES de Berkeley, possuindo o código fonte aberto. Fornece suporte às linguagens SQL92/SQL99, além de outras funcionalidades avançadas. O POSTGRES foi pioneiro em muitos conceitos objeto-relacionais que agora estão se tornando disponíveis em alguns bancos de dados comerciais.

Os SGBD relacionais tradicionais suportam um modelo de dados composto por uma coleção de relações com nome, contendo atributos de um tipo específico. Nos sistemas comerciais em uso, os tipos possíveis incluem número de ponto flutuante, inteiro, cadeia de caracteres, valores monetários e data. É amplamente reconhecido que este modelo não é adequado para as futuras aplicações de processamento de dados, devido a crescente necessidade por manipulação de dados complexos.

O modelo relacional substituiu com sucesso os modelos anteriores principalmente pela sua simplicidade. Entretanto, esta simplicidade tornou a implementação de certas aplicações muito difícil. O PostgreSQL oferece recursos adicionais pela incorporação de

⁶Documentação extraída do site oficial do PostgreSQL, disponível em <http://www.postgresql.org>.

vários conceitos e funcionalidades. São estas funcionalidades que colocam o PostgreSQL dentro da categoria de banco de dados referida como objeto-relacional.

Vale ressaltar que existe uma diferença entre os SGBD objeto-relacionais daqueles referidos como puramente orientados a objetos que, em geral, não são muito adequados para apoiar as linguagens tradicionais de banco de dados relacional. Portanto, embora o PostgreSQL possua algumas funcionalidades de orientação a objetos, está firmemente ligado ao mundo dos bancos de dados relacionais. Na verdade, alguns bancos de dados comerciais incorporaram recentemente funcionalidades nas quais o PostgreSQL foi pioneiro.

Não só por estes motivos, mas também pelo fato do PostgreSQL ser atualmente considerado por diversos autores na área de bancos de dados como um sistema eficiente e robusto, que sua utilização neste trabalho justifica-se.

3 METODOLOGIA

Este estudo bibliométrico, realizado por meio da análise das características das fontes de informação expressas nas citações de teses, foi construído a partir de teses apresentadas ao Programa de Pós-Graduação em Administração (PPGA) da Escola de Administração (EA) da Universidade Federal do Rio Grande do Sul (UFRGS).

Para o desenvolvimento deste trabalho, foi utilizada a pesquisa bibliográfica e foi aplicado o processo de construção em bancos de dados, baseado na arquitetura em três esquemas: do modelo conceitual, passando pelo modelo físico, até o modelo de implementação. Com o objetivo final de realizar a análise de citações das teses.

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

A pesquisa bibliográfica deu base a para a aquisição de conhecimento acerca dos temas envolvidos no projeto, principalmente no que diz respeito a bibliometria e análise de citações. Envolveu, basicamente, consultas a livros de referência, teses e artigos científicos.

Para a realização desta investigação, os dados foram coletados no catálogo *on-line* (*Online Public Access Catalogue – OPAC*) do Sistema de Bibliotecas da UFRGS, utilizando o Sistema de Automação de Bibliotecas (SABi). Mais precisamente, o catálogo de teses e dissertações pertencentes à Biblioteca Setorial da EA, selecionando as teses disponibilizadas em texto completo, defendidas de 1999 a 2006 e orientadas por professores pertencentes ao Departamento de Ciências Administrativas (DCA), representando até esta data um total de 80 teses.

Os dados encontram-se disponíveis mediante pesquisa no *site* do SABi⁷, selecionando o ícone catálogo, tipo de material, teses e dissertações. Selecionado o

⁷ <http://sabix.ufrgs.br/ALEPH?415825923>.

catálogo específico, a seguinte linguagem de comando – expressão de busca – permite recuperar as teses destacadas por meio da Pesquisa CCL (*Commum Comand Language*):

WBP = adm and WUN = adm01 and WPI = pd and WYR = 1999 -> 2006 and WLI = texto complete

Onde: WBP para biblioteca setorial, WUN para produção intelectual do DCA, WPI para selecionar somente as teses, WYR para delimitar o intervalo de tempo a ser estudado, e WLI para encontrar as tese que estão disponíveis em texto completo no catálogo.

Todas as teses disponíveis no SABi têm seus arquivos passíveis de importação (*download*) e têm por padrão sua conversão para arquivos com extensão *.pdf e são disponibilizadas como imagem pelo *software Adobe*. Para trabalhar os dados textuais extraídos de arquivos em formato pdf, foram utilizadas as ferramenta de conversão *xpdf* do linux usando seu utilitário *pdftolinux* e *Solid Converter PDF* do Windows, passando estes arquivos para *.txt.

Para as etapas de criação do banco de dados foram desenvolvidos: o modelo conceitual, utilizando o modelo Entidade-Relacionamento (ER); o mapeamento deste para o modelo físico; e finalmente o modelo de implementação, utilizando o modelo relacional.

O resultado da modelagem conceitual é o diagrama ER, que pode ser observado na Figura 3.1.

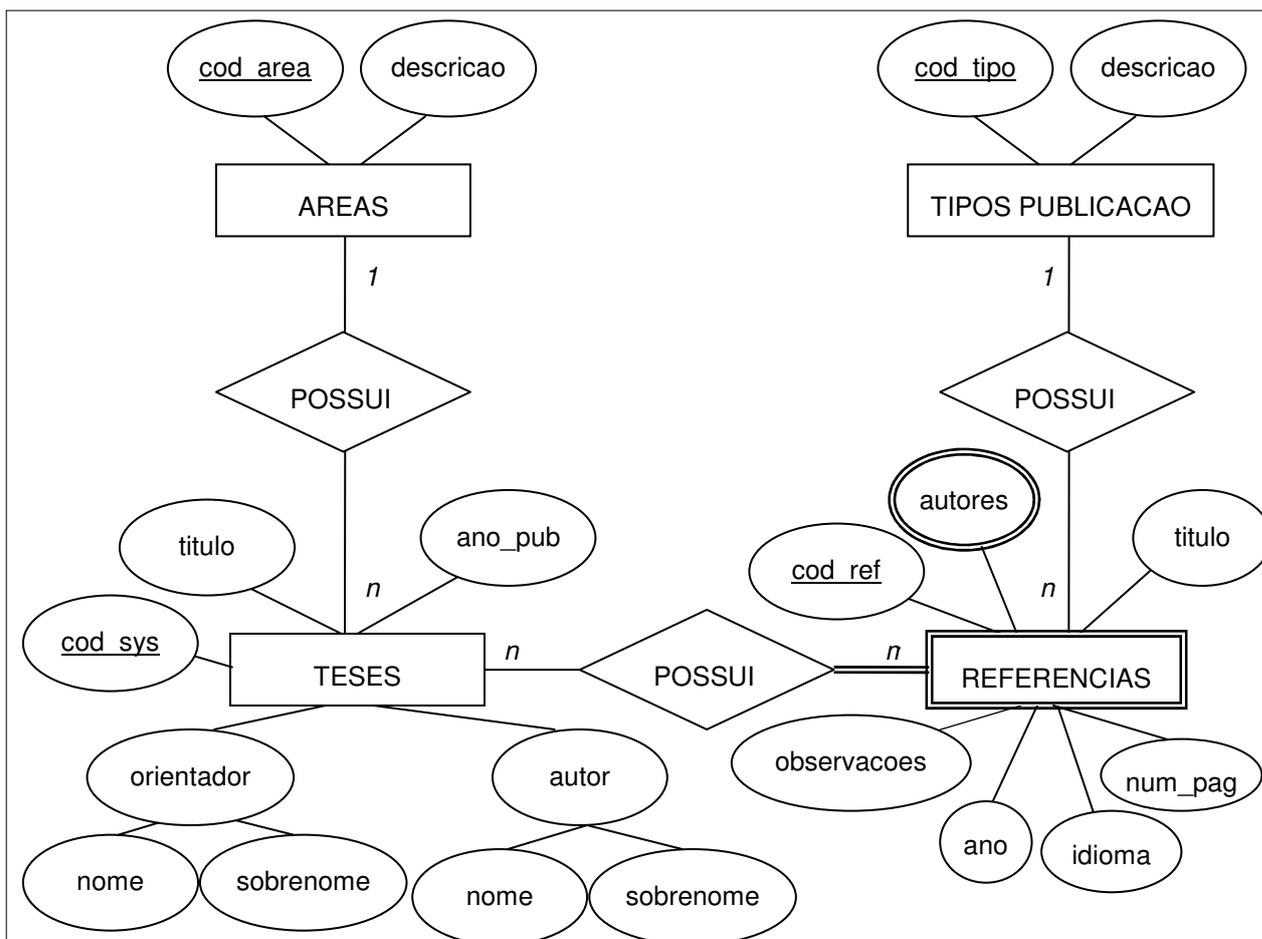


Figura 3.1: Modelo ER do banco de dados de teses.

Fonte: Elaborada pelo autor.

Este modelo é resultante das seguintes características do banco de dados de teses:

1. A entidade AREAS representa as áreas existentes no PPPGA/EA/UFRGS, onde apenas a informação que interessa a este trabalho é aquela relativa à área na qual as teses estão inseridas. Neste caso, cada tese possui uma área específica e, por outro lado, uma área específica pode possuir várias teses. Esta característica está representada no modelo pela cardinalidade “1:n” no relacionamento POSSUI, entre as entidades AREAS e TESES.
2. A entidade TIPOS_PUBLICACAO representa os possíveis tipos das publicações presentes nas referências das teses, podendo ser por exemplo: livro, periódico, artigo, capítulo de livro, dentre outros. A informação que interessa a este trabalho é qual é o tipo de publicação específico de cada referência. Neste caso, cada referência é de um tipo específico de publicação e, por outro lado, um tipo de publicação pode ser de várias referências. Esta característica está representada no

modelo pela cardinalidade “1:n” no relacionamento POSSUI, entre as entidades TIPOS_PUBLICACAO e REFERENCIAS.

3. A entidade TESES representa as teses existentes no PPPGA/EA/UFRGS, entre os anos de 1999 e 2006, onde as informações que interessam a este trabalho são: o código de cada tese no sistema de bibliotecas da EA/UFRGS, o título da tese, o nome do autor (composto por sobrenome e nome), o nome do orientador (composto por sobrenome e nome), o ano de publicação da tese e sua relação com as áreas de pesquisa.
4. A entidade REFERENCIAS representa todas as referências citadas em todas as teses, onde as informações que interessam a este trabalho são: um código para identificar cada referência, o título da obra, os nomes dos autores (pode ser mais de um), o idioma, número de páginas e o ano de publicação da obra, e algumas observações que podem variar de acordo com o tipo de obra (por exemplo, para artigos a revista, o local, o mês etc). Esta é uma entidade fraca porque a existência de uma referência no banco de dados depende fortemente da existência de uma tese que a contenha, ou seja, não faz sentido ter uma referência no banco de dados, que não pertença a uma tese.
5. Existe um forte relacionamento POSSUI entre as entidades TESES e REFERENCIAS, com a cardinalidade “n:m”, que representa a seguinte participação: uma tese pode possuir várias referências e uma referência pode ser citada em várias teses.

O mapeamento do modelo ER para o modelo relacional foi realizado utilizando a ferramenta de projeto DBDesigner, que é uma ferramenta gratuita e de grande qualidade. A utilização desta ferramenta tem a finalidade de facilitar o trabalho do projetista de banco de dados, fornecendo uma melhor visão do modelo conceitual e como ele será posteriormente implementado. O resultado do mapeamento entre o modelo conceitual (ER) e o modelo de implementação (relacional), pode ser observado na Figura 3.2.

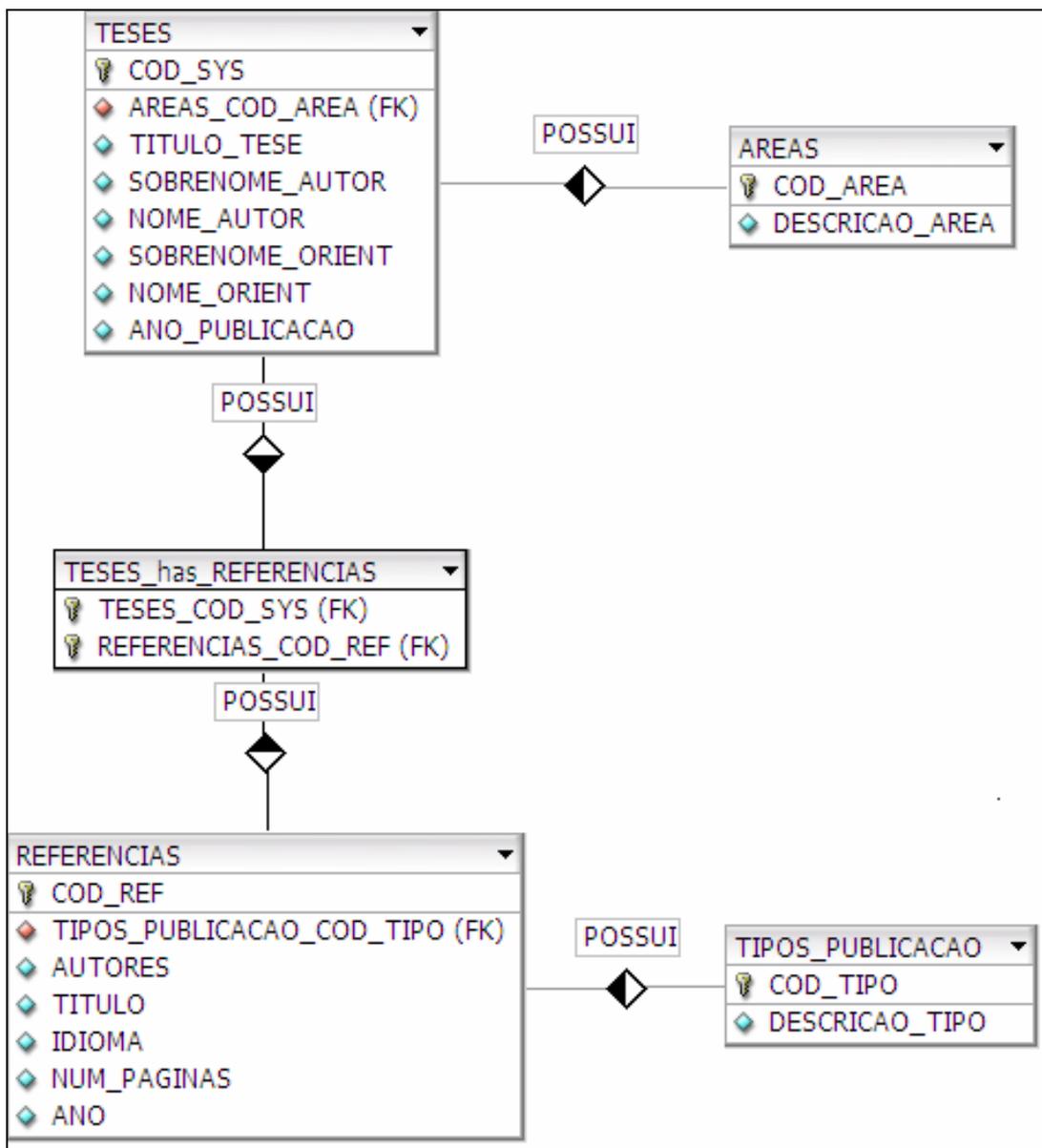


Figura 3.2: Mapeamento modelo ER-relacional feita no DBDesigner.
 Fonte: Elaborada pelo autor.

O modelo relacional, foi implementado utilizando a linguagem SQL no SGBD PostgreSQL. A escolha por este SGBD baseou-se nas características, que conforme confirma Date (2005), o mesmo possui: *software* de código aberto (escolha totalmente gratuita e livre); capacidade de lidar com grande volume de dados (existem aplicações com tabelas de 200 milhões de linhas executando com resultados excelentes), multiplataforma (executa nos principais sistemas operacionais tipo Unix, incluindo Linux e Mac OS X e em Windows); grande número de interfaces de programação (ODBC, Java, TCL/TK, PHP, Perl, Python, dentre outras); e inúmeras funcionalidades técnicas, dentre outras:

integridade referencial, número ilimitado de linhas e índices em tabelas, uso otimizado de recursos do Sistema Operacional, *triggers*, *views*, *stored procedures*, *joins* etc.

O *script* de criação do banco de dados de teses, que descreve todos os comandos na ordem em que foram executados, é apresentado na Figura 3.3.

```

CREATE TABLE AREAS (
    COD_AREA          CHAR(4)          NOT NULL,
    DESCRICAO_AREA   VARCHAR(100)     NOT NULL,
    PRIMARY KEY (COD_AREA));

CREATE TABLE TESES (
    COD_SYS           CHAR(6)          NOT NULL,
    COD_AREA          CHAR(4)          NOT NULL,
    TITULO_TESE      VARCHAR(300)     NOT NULL,
    SOBRENOME_AUTOR  VARCHAR(50)      NOT NULL,
    NOME_AUTOR       VARCHAR(100)     NOT NULL,
    SOBRENOME_ORIENT VARCHAR(50)      NOT NULL,
    NOME_ORIENT      VARCHAR(100)     NOT NULL,
    ANO_PUBLICACAO   DATE              NOT NULL,
    PRIMARY KEY (COD_SYS),
    FOREIGN KEY (COD_AREA) REFERENCES AREAS (COD_AREA));

CREATE TABLE TIPOS_PUBLICACAO (
    COD_TIPO          CHAR(4)          NOT NULL,
    DESCRICAO_TIPO   VARCHAR(100)     NOT NULL,
    PRIMARY KEY (COD_TIPO));

CREATE TABLE REFERENCIAS (
    COD_REF           INTEGER          NOT NULL,
    COD_TIPO          CHAR(4)          NOT NULL,
    AUTORES           VARCHAR(300)     NOT NULL,
    TITULO            VARCHAR(200)     NOT NULL,
    IDIOMA            VARCHAR(50)      NOT NULL,
    NUM_PAGINAS       INTEGER,
    ANO               CHAR(4)          NOT NULL,
    PRIMARY KEY (COD_REF),
    FOREIGN KEY (COD_TIPO) REFERENCES TIPOS_PUBLICACAO (COD_TIPO));

CREATE TABLE REFERENCIAS_TESSES (
    COD_SYS           CHAR(6)          NOT NULL,
    COD_REF           INTEGER          NOT NULL,
    PRIMARY KEY (COD_SYS, COD_REF),
    FOREIGN KEY (COD_SYS) REFERENCES TESES (COD_SYS),
    FOREIGN KEY (COD_REF) REFERENCES REFERENCIAS (COD_REF));

```

Figura 3.3: *Script* de criação do banco de dados de teses, na linguagem SQL.

Fonte: Elaborada pelo autor.

Uma vez criado o banco de dados, o mesmo foi populado e consultas foram realizadas. O resultado obtido após o banco de dados ter sido populado foi o seguinte:

- As áreas existentes no PPPGA/EA/UFRGS são seis: Gestão da Tecnologia e da

Produção, Sistemas de Informação e de Apoio à Decisão, Contabilidade e Finanças, Marketing, Recursos Humanos, e Organizações.

- As teses defendidas no PPGA/EA/UFRGS entre os anos de 1999 e 2006, que estão disponíveis *on-line* são 80 (oitenta).
- Somadas todas as referências presentes nestas teses, obteve-se um total de 12.557. Estes são os dados que serão tratados e mais aprofundados nas análises.
- Os tipos de publicação existentes entre as referências foram divididos em seis grupos: livros, capítulos de livros, artigos, anais de eventos, teses e dissertações.

Nesta etapa de população do banco de dados diversas dificuldades foram encontradas, dentre as quais vale ressaltar: a falta de utilização das normas para fazer as referências muitas vezes impossibilitou a classificação da obra em um tipo de publicação; algumas vezes observou-se ausência de algumas informações importantes nas referências colocadas; e outros problemas de má elaboração e falta de padronização das referências. Diante das impossibilidades de classificação das referências, algumas foram descartadas e no total o número de referências que serviram de base para as análises foi de 11.358.

Sendo assim, a inserção das tuplas no banco de dados foi feita manualmente. Cada uma das referências teve que ser examinada, classificada e inserida, devido a complexidade de seguir as normas de criação de referências, que poderiam gerar erros, caso a inserção fosse feita automaticamente. Por exemplo, um erro poderia ser gerado devido a norma que estabelece que o mesmo autor quando referenciado mais de uma vez, a partir da segunda vez, em vez de ser colocado seu sobrenome, deve ser colocado um traço. Neste caso, porém, no banco de dados este traço tem que ser substituído pelo sobrenome do autor.

Com as consultas feitas sobre o banco de dados, tendo por base as citações expressas nas teses, pretende-se analisar os seguintes elementos:

- a) tipologia dos materiais bibliográficos empregados nas citações, objetivando melhorar a distribuição dos recursos financeiros disponíveis para investimento em inclusão, substituição e prioridades na coleção geral da biblioteca;
- b) obsolescência dos materiais bibliográficos citados, visando traçar diretrizes para o armazenamento retrospectivo da coleção;
- c) títulos de periódicos citados, para determinar a coleção-núcleo de periódicos e preparar possíveis cortes na coleção de periódicos devido a, por exemplo, eventuais reduções no orçamento da biblioteca;

d) idioma dos materiais consultados, identificando os idiomas que devem ocorrer nos materiais bibliográficos selecionados em futuras aquisições por compra, permuta ou doação.

4 RESULTADOS E ANÁLISES

Antes de iniciar a análise de citações foram realizadas algumas consultas que descrevem com maiores detalhes as 80 teses estudadas. Alguns destes resultados iniciais podem ser observados a seguir.

A Figura 4.1 apresenta a distribuição das teses por área de pesquisa do PPGA/EA/UFRGS. Por este gráfico pode-se observar claramente que algumas áreas se destacam no programa, em detrimento de outras. A área de Gestão da Tecnologia e da Produção é a responsável pelo maior número (23 teses) e, por outro lado, a área de Contabilidade e Finanças com o menor número (3 teses).

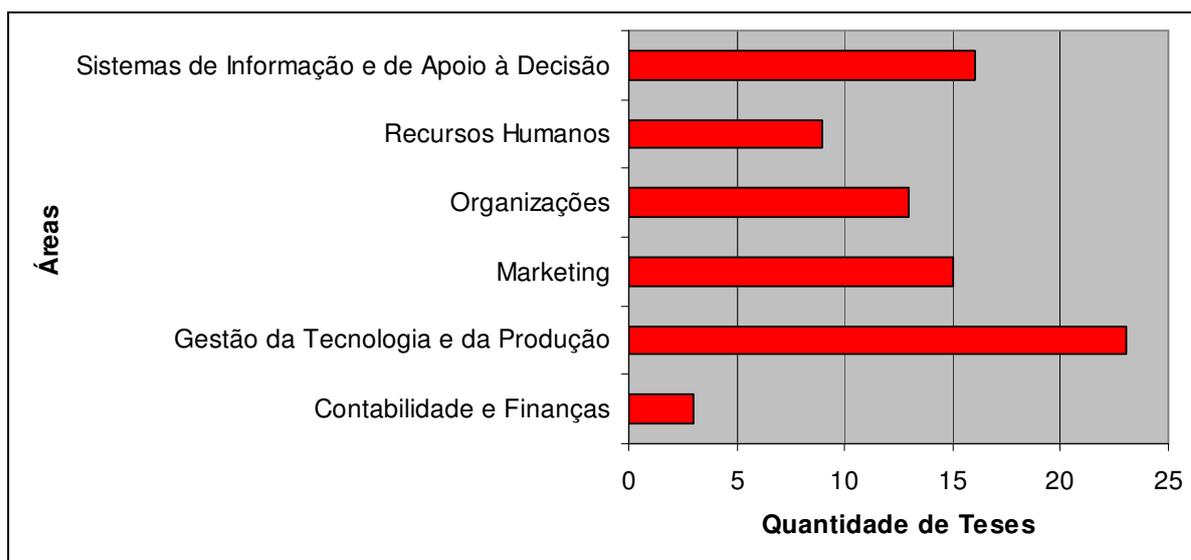


Figura 4.1: Gráfico da relação de quantidade de teses por área.
Fonte: Elaborada pelo autor.

A Figura 4.2 apresenta um gráfico que ilustra a distribuição da quantidade de orientações pelos professores. Esta distribuição apresenta-se bastante equilibrada, uma vez que a maioria dos professores encontra-se na média de 3 orientações. O maior número de orientações foi de 8 para um professor e 4 professores orientaram apenas 1. Total de 26 professores.

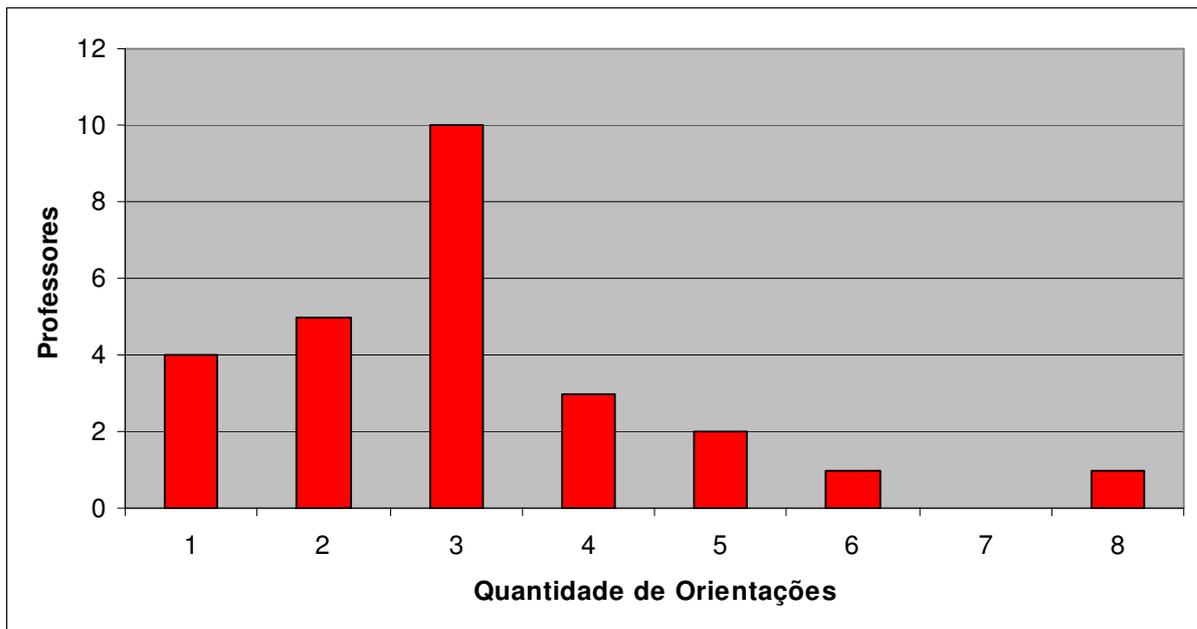


Figura 4.2: Gráfico da relação de quantidade de orientações por professores.
 Fonte: Elaborada pelo autor.

A Figura 4.3 apresenta um gráfico da distribuição por ano da publicação de teses. Pode-se observar que o ano quando houve o maior número de publicação de teses foi em 2005 com 20 teses, o ano em que houve o menor número de teses publicadas foi em 2003 com apenas 3, sendo que a média de teses defendidas por ano é de 10 teses.

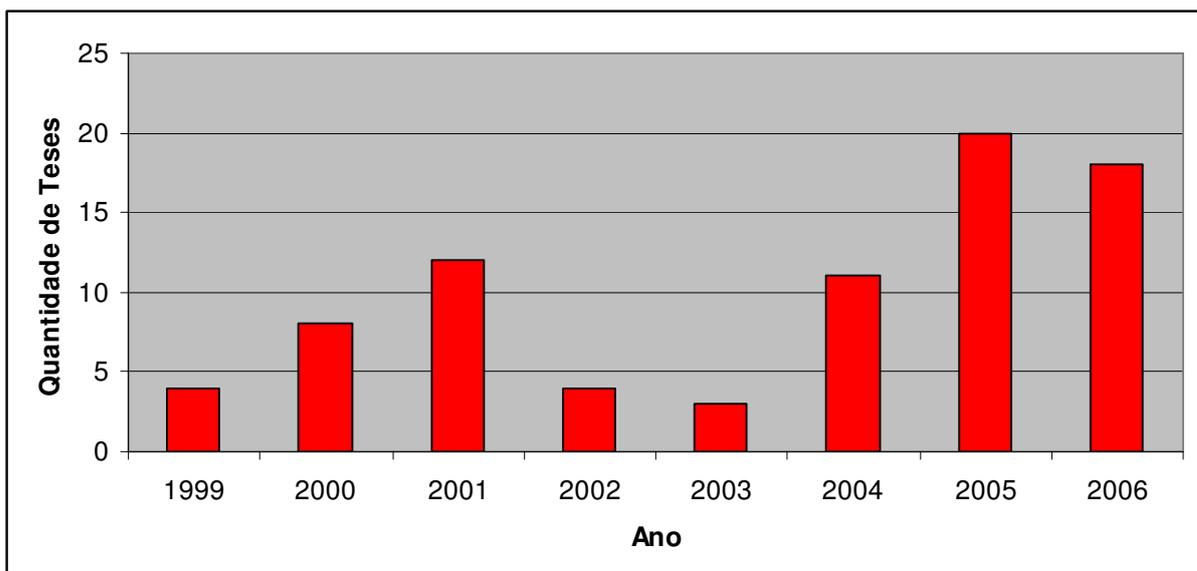


Figura 4.3: Gráfico da relação de quantidade de teses publicadas por ano.
 Fonte: Elaborada pelo autor.

Inicialmente a quantidade de referências contidas nas teses era de 12.557, porém devido a erros encontrados que foram cometidos na elaboração das mesmas, este número foi reduzido em 9,5%, resultando num conjunto de 11.358 referências para serem estudadas.

Pode-se observar que não há uma distribuição equilibrada do número de referências por tese. A média do número de referências em cada tese é de 157, mas a maioria 53% das teses está muito abaixo da média com aproximadamente 110 referências por tese. Também há dois casos muito fora do padrão: uma tese com 333 referências e a tese que apresentou o maior número de referências por tese (452), ambas da área de Recursos Humanos. Há também o caso da tese que contém o menor número de referências, 32, na área de Sistemas de Informação e de Apoio à Decisão.

A análise de citações proposta como objetivo desta pesquisa visa responder à tipologia dos materiais bibliográficos empregados nas citações; a obsolescência dos materiais bibliográficos citados; os títulos de periódicos citados; e os idiomas dos materiais consultados. Diversas consultas foram realizadas no banco de dados para alcançar tais objetivos. Por motivos de melhor apresentação, optou-se por colocar todos os comandos SQL utilizados na fase de consultas ao banco de dados e posterior análise no Anexo A.

4.1 Análise da tipologia das referências

De acordo com o que pôde ser observado e classificado pelo tipo de publicação, dentre as referências citadas nas teses, observa-se a distribuição conforme a Tabela 4.1.

Tabela 4.1: Distribuição das referências por tipos de publicação.

Tipo de publicação	Quantidade	Percentual
Livros	7141	68,9
Artigos	2267	21,8
Anais de eventos	659	6,3
Teses	133	1,2
Dissertações	120	1,1
Capítulos de livro	34	0,3

Fonte: Elaborada pelo autor.

A Figura 4.4 apresenta uma melhor forma de observar os mesmos dados através de um gráfico da distribuição das referências por tipos de publicação.

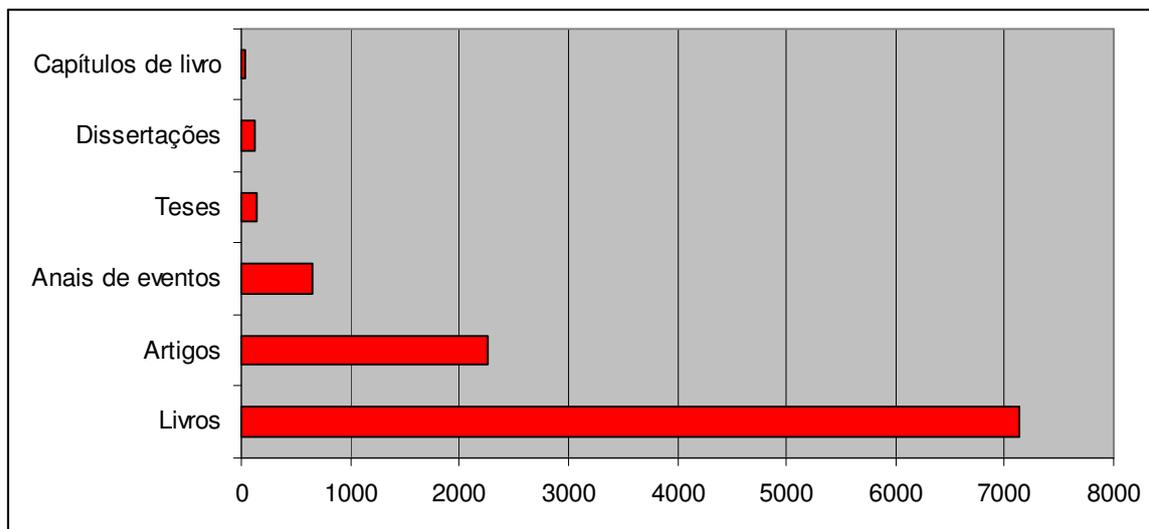


Figura 4.4: Gráfico da distribuição das referências por tipos de publicação.

Fonte: Elaborada pelo autor.

Diante destes primeiros resultados é interessante notar que a grande maioria das referências utilizadas para elaboração das teses é extraída de livros. Vale ressaltar que nem todas as obras puderam ser seguramente classificadas em algum tipo de publicação devido a ausência de informações contidas na própria referência. As obras que não entraram nesta análise correspondem a aproximadamente 8,8% .

Outra consulta foi realizada visando levantar a quantidade de referências que foi extraída da internet e observou-se que este é um número bastante pequeno. Apenas em 429 referências constavam os endereços de acesso a algum material da internet.

4.2 Análise da obsolescência das referências

Com relação ao segundo ponto da análise de citações, a obsolescência dos materiais bibliográficos, pôde-se observar que a grande maioria das obras data das duas últimas décadas. A Tabela 4.2 mostra a distribuição das obras por décadas.

Tabela 4.2: distribuição das obras por décadas.

Década	Quantidade	Percentual
Até 1959	112	1,02
Entre 1960 e 1969	237	2,16
Entre 1970 e 1979	660	6,02
Entre 1980 e 1989	1792	16,36
Entre 1990 e 1999	6067	55,41
De 2000 a 2006	2081	19,00

Fonte: Elaborada pelo autor.

Diante destas informações e levando-se em conta que a grande área do programa de pós-graduação é a administração, observa-se que existe uma grande probabilidade que a maioria das obras utilizadas como referências nas teses não estejam obsoletas. Porém melhores análises e conclusões poderão ser feitas por especialistas nesta área.

Uma vez que as décadas mais citadas nas teses, como observado no resultado anterior, são as duas últimas, optou-se por apresentar uma visão mais detalhada como pode ser observado no gráfico da distribuição de obras por ano da Figura 4.5.

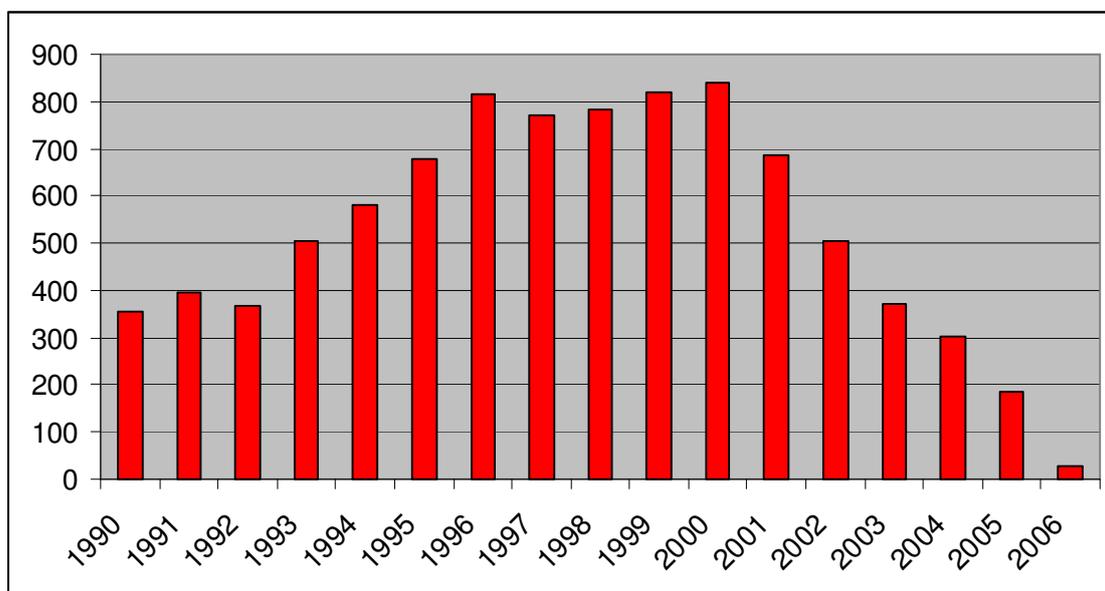


Figura 4.5: Gráfico da distribuição de referências entre os anos de 1990 e 2006.

Fonte: Elaborada pelo autor.

Por este gráfico observa-se que há uma grande utilidade de obras de meados da década de 90 até o início dos anos 2000. Com estas informações disponíveis aos administradores da biblioteca, torna-se mais fácil a tarefa de definir quando as obras devem passar a ser descartadas da coleção.

4.3 Análise dos títulos de periódicos mais citados

A lista completa de periódicos é enorme. No total foram encontrados mais de 2.800 referências a periódicos nacionais e internacionais, porém dentre estes alguns são citados com uma frequência maior.

Dentre os periódicos nacionais destacam-se a “Revista de Administração” que foi a mais citada com 25% do total de periódicos nacionais, seguida da “Revista de Administração de Empresas” com 11%. É interessante notar que uma revista não científica tenha sido bastante referenciada, a “Revista Exame” com 8% de citações do total de periódicos nacionais. Outra informação interessante é que aproximadamente 8% dos periódicos nacionais citados pertencem a outras áreas, tais como saúde, agronegócio, engenharia etc.

Dentre os periódicos internacionais o mais citado foi o “*American Journal of Sociology*” correspondendo a 21% das referências, seguido do “*Academy of Management Journal*” com 16%, e do “*The Journal of Business Strategy*” com 10%. Também dentre as publicações internacionais observou-se a quantidade de periódicos de outras áreas, que, neste caso, correspondem a 6%.

Acredita-se que, uma vez que o responsável pelo acervo da biblioteca tenha estas informações, a tarefa de decidir quais periódicos devem ser adquiridos entre tantos, será mais fácil.

4.4 Análise dos idiomas das referências

Por fim, verifica-se como ocorre a classificação das obras utilizadas como referências nas teses pelos idiomas. A Figura 4.6 mostra um gráfico com esta classificação.

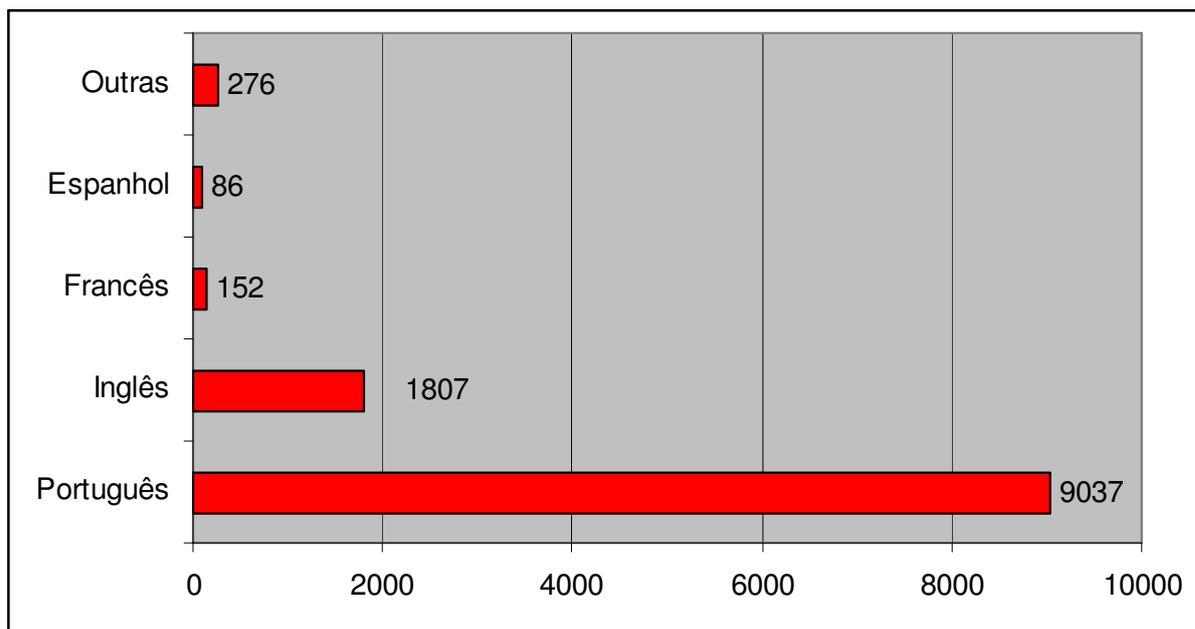


Figura 4.6: Gráfico da distribuição das referências pelos idiomas.
Fonte: Elaborada pelo autor.

Pode-se observar que o português é o idioma largamente mais utilizado, seguido do idioma inglês. É interessante notar que, apesar da proximidade dos idiomas português e espanhol, este último aparece com uma frequência baixa, inferior a quantidade de obras em francês. Acredita-se que esta informação também seja útil para a tomada de decisão dos administradores da biblioteca, no que diz respeito ao desenvolvimento de coleções.

5 CONCLUSÕES

Realizou-se o estudo do uso da informação técnica e científica, por meio da análise das citações utilizadas na elaboração de teses acadêmicas. Foram analisados o conjunto de dados extraídos das teses defendidas entre os anos de 1999 a 2006, orientadas por professores pertencentes ao Departamento de Ciências Administrativas da Escola de Administração da Universidade Federal do Rio Grande do Sul (UFRGS), teses disponibilizadas em texto completo no catálogo *on-line* (*Online Public Access Catalogue - OPAC*) do Sistema de Bibliotecas da UFRGS.

O objetivo geral foi de criar um banco de dados das referências bibliográficas de teses de um programa de pós-graduação e posteriormente estudar o uso da informação técnica e científica por meio da análise das citações. Este objetivo foi alcançado e o banco de dados está disponível para realizar outras consultas que poderão surgir.

Os objetivos específicos foram de coletar as teses disponíveis em formato digital, que serviram de base para a análise; modelar e desenvolver um banco de dados das referências bibliográficas utilizadas em todas as teses; e criar consultas no banco de dados permitindo analisar a tipologia dos materiais bibliográficos utilizados nas citações, sua obsolescência, os títulos de periódicos citados e o idioma dos materiais consultados. Os objetivos específicos também foram atingidos.

A coleta das teses, que serviram de base para a análise, resultou num número total de 80 teses que correspondem à totalidade do conjunto de teses disponíveis no sistema *on-line*. Nesta etapa, a maior dificuldade encontrada foi a transformação uma a uma das teses do formato em que se encontravam (*.pdf) para o formato em que foi possível manipular as suas informações (*.txt). Foi um trabalho exaustivo, porém necessário uma vez que pretendeu-se garantir a preservação as informações da forma em que estavam disponibilizadas.

Na etapa de modelagem e desenvolvimento do banco de dados das referências bibliográficas utilizadas em todas as teses, foi realizada inicialmente as modelagens conceitual e física, para posterior implementação no modelo relacional. A modelagem conceitual foi feita no modelo Entidade-Relacionamento. Para a implementação do banco de dados foi utilizada a linguagem SQL, no sistema gerenciador de bancos de dados PostgreSQL.

Sem dúvida a etapa desta pesquisa que consumiu mais tempo e trabalho foi a de popular o banco de dados. Cada uma das 11.358 referências bibliográficas foram analisadas e classificadas antes de serem inseridas no banco de dados uma a uma. Também são desta fase os maiores problemas enfrentados na realização do trabalho, devido principalmente à falta de padronização dos dados, gerando, muitas vezes, inconsistências que tiveram que ser tratadas. Acredita-se que um melhor uso das normas de escrita de referências bibliográficas poderia ter melhorado a qualidade dos resultados obtidos nesta etapa do trabalho.

A última etapa da pesquisa de criar consultas no banco de dados e fazer a análise dos resultados obtidos foi realizada satisfatoriamente. As consultas foram relativamente simples, porém apresentaram os resultados desejados. Acredita-se que as informações obtidas destas consultas, se passadas a pessoas especializadas na área de biblioteconomia, poderão ser melhor analisadas.

Como resultados das análises de citações realizadas nesta pesquisa, os resultados mais expressivos encontrados estão colocados a seguir.

A área de Gestão da Tecnologia e da Produção é a responsável pelo maior número de teses desenvolvidas no período.

A distribuição da quantidade de orientações pelos professores apresenta-se equilibrada, uma vez que a maioria dos professores encontra-se dentro da média.

Não há uma distribuição equilibrada do número de referências por tese. A média número de referências em cada tese é de 157, mas a maioria 53% das teses está abaixo da média com aproximadamente 110 referências por tese.

Com relação aos tipos de publicação utilizadas, a grande maioria das referências para elaboração das teses é extraída de livros, seguido de artigos.

A obsolescência dos materiais bibliográficos não foi observada, uma vez que a grande maioria das obras data das duas últimas décadas. Porém melhores análises e conclusões poderão ser feitas por especialistas nesta área.

A lista completa de periódicos utilizados como referência é enorme. No total foram encontrados mais de 2.800 referências a periódicos nacionais e internacionais, e esta lista completa está disponível no banco de dados. Acredita-se que, uma vez que o responsável pelo acervo da biblioteca tenha estas informações, a tarefa de decidir quais periódicos devem ser adquiridos entre tantos, será mais fácil.

Por fim, o português é o idioma largamente mais utilizado, seguido do idioma

inglês, francês e espanhol.

Acredita-se que estas informações sejam úteis para a tomada de decisão dos administradores da biblioteca, no que diz respeito ao desenvolvimento de coleções. Também acredita-se que uma melhor análise possa ser feita utilizando os dados do banco de dados, por especialistas na área, que possam contribuir para melhorar a distribuição dos recursos financeiros disponíveis para investimento em material informacional; para traçar diretrizes para o armazenamento retrospectivo da coleção; para determinar a composição da coleção de periódicos; para avaliar em que idiomas devem ser selecionados os materiais bibliográficos para futuras aquisições por compra, permuta ou doação; bem como para preparar possíveis cortes na coleção devido a eventuais reduções no orçamento da biblioteca.

Como sugestões para trabalhos futuros, que aproveitem os resultados deste, sugere-se a real utilização destes resultados por parte dos administradores da biblioteca, alcançando as contribuições enunciadas. Também sugere-se uma maior exploração dos dados que estão disponíveis no banco de dados, uma vez que estes contém uma riqueza de informações ainda inexplorada. Por fim, sugere-se a realização de trabalho semelhante a este em outras bibliotecas.

6 ANEXOS

6.1 Anexo A – Consultas ao BD usando a linguagem SQL

Consulta 1: Mostrar as quantidades de teses classificadas por áreas

```
SELECT DESCRICAO_AREA, COUNT (COD_REF)
FROM AREAS, TESES
WHERE AREAS.COD_AREA = TESES.COD_AREA
GROUP BY COD_AREA
ORDER BY 2 DESC;
```

Consulta 2: Mostrar os orientadores e as quantidades de teses orientadas por eles

```
SELECT COUNT (*), NOME_ORIENT
FROM TESES
GROUP BY NOME_ORIENT
ORDER BY 2;
```

Consulta 3: Mostrar todas as referências ordenadas por tipo de publicação:

```
SELECT TIPOS_PUBLICACAO.DESCRICAO_TIPO, REFERENCIAS.TITULO
FROM REFERENCIAS, TIPOS_PUBLICACAO
WHERE REFERENCIAS.COD_TIPO = TIPOS_PUBLICACAO.COD_TIPO
ORDER BY 1;
```

Consulta 4: Mostrar a quantidade de referências de cada tipo de publicação:

```
SELECT COUNT (*), COD_TIPO
FROM REFERENCIAS
GROUP BY COD_TIPO;
```

Consulta 5: Mostrar a quantidade de referências agrupadas por idioma.

```
SELECT COUNT (*), IDIOMA  
FROM REFERENCIAS  
GROUP BY IDIOMAS;
```

Consulta 6: Mostrar as referências ordenadas pelo ano de publicação.

```
SELECT TITULO, ANO  
FROM REFERENCIAS  
ORDER BY ANO DESC;
```

7 REFERÊNCIAS BIBLIOGRÁFICAS

ACKERSON, L. G. Is age an appropriate criterion for moving journals to storage? **Collection Management**: devoted to the management of library collections, New York, v. 26, n. 3, p. 63-76, 2001.

BROADUS, R. N. The applications of citation analyses to library collection building. *In*: VOIGT, M. J.; HARRIS, M. H. (Editors). **Advances in librarianship**, New York: Academic Press, 1977. v. 7, p. 299-335.

CHOO, C. W. **A organização do conhecimento**: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões. São Paulo: SENAC, 2003.

CODD, E.F.. **A Relational Model of Data for Large Shared Data Banks**. *In*: *Communications of the ACM*. v. 6, 1970.

CÔRTEZ, P. L. A importância da literatura cinzenta disponível na Internet para as áreas de Ciências Contábeis e Administração de Empresas. *Revista Brasileira de Gestão de Negócios*, São Paulo, v. 8, n. 20, p. 13-22, jan./abr. 2006.

DATE, C. J. **Introdução a Sistemas de Banco de Dados**. Rio de Janeiro: Campus, 2005.

DEVIN, R. B.; KELLOGG, M. The serial/monograph ratio in research libraries: budgeting in light of citation studies. **College and Research Libraries**, Chicago, v. 51, n. 1, p. 46-54, Jan. 1990.

DIAS, M. M. K.; PIRES, D. **Formação e desenvolvimento de coleções de serviços de informação**. São Carlos: EdUFSCar, 2003.

DOU, H. A bibliometria e os softwares *Matheo Analyzer* e *Matheo Patent*. *In*: TARAPANOFF, K. (Org.). **Inteligência, informação e conhecimento**. Brasília, DF: UNESCO/IBICT, 2006. p. 339-378.

ELMASRI, R.; NAVATHE, S. B. **Sistemas de Banco de Dados : fundamentos e aplicações**. 3 ed. Rio de Janeiro : LTC - Livros Técnicos e Científicos, 2005. 781 p.

FIGUEIREDO, N. M. de; LIMA, R. C. M. de. Seleção e aquisição: da visão clássica à moderna aplicação de técnicas bibliométricas. *In*: FIGUEIREDO, N. M. de. **Desenvolvimento & avaliação de coleções**. 2. ed. ver. atual. Brasília, DF: Thesaurus, 1998. p. 67-96.

GARCIA-MOLINA, H. ; ULLMAN, J. D.; WIDOM, J. **Implementação de Sistemas de Banco de Dados**. Rio de Janeiro: Ed. Campus, 2002. 697 p.

GARFIELD, E. Is citation frequency a valid criterion for selecting journals? **Essays of an Information Scientist**, Philadelphia, v. 1, p. 289-290, 1977.

GREENE, R. J. Computer analysis of local citation information in collection management. **Collection Management**: devoted to the management of library collections, New York, v. 17, n. 4, p. 11-24, June 1993.

GOEBEL, M.; GRUENWALD, L. **A survey of data mining and knowledge discovery software tools**. ACM SIGKDD Explorations, v.1, n.1, June, 1999.

IRVINE, J.; MARTIN, B. R. **International comparisons of scientific performance revisited**. Scientometrics, Amsterdam, v. 15, n. 5-6, p. 369-392, 1989.

KLAES, R. R. Sistema de informação gerencial para desenvolvimento de coleções. **Ciência da Informação**, Brasília, DF, v. 20, n. 2. p. 220-228, jul./dez. 1991.

KNIEVEL, J. E.; KELLSEY, C. Citation analysis for collection development: a comparative study of eight humanities fields. **The Library Quarterly**, Chicago, v. 75, n. 2, p. 142-168, Apr. 2005.

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

LE COADIC, Y. **A ciência da informação**. 2. ed. Brasília, DF: Briquet de Lemos, 2004.

LETA, J.; CRUZ, C. H. de B. A produção científica brasileira. *In*: VIOTTI, E. B.; MACEDO, M. de M. **Indicadores de ciência, tecnologia e inovação no Brasil**. Campinas: Editora Unicamp, 2003. p. 121-168.

LINE, Maurice B. Rank lists based on citations and library uses as indicators of journal usage in individual libraries. **Collection Management**: devoted to the management of library collections, New York, v. 2, n. 4, p. 313-316, 1978.

MACIAS-CHAPULA, C. A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ciência da Informação**, Brasília, DF, v. 27, n. 2, p. 141-148, maio/ago. 1998. Disponível em: <http://www.scielo.br/pdf/ci/v27n2/macias.pdf>. Acesso em: 07 jan 2007 Tradução do inglês de Irati Antonio.

MEADOWS, A. J. **A comunicação científica**. Brasília, DF: Briquet de Lemos, 1999.

NEGRETE GUTIÉRREZ, M. C. El papel del bibliotecólogo em el desarrollo de colecciones em la biblioteca universitária. **Investigación Bibliotecológica**: archivonomía, bibliotecología e información, México, v. 7, n. 14, p. 5-7, ene./jun. 1993.

NEVES, D. L. F. **PostgreSQL: Conceitos e Aplicações**. São Paulo: Érica, 2002.

_____. La comunidad universitária y el desarrollo de colecciones. *In*: NEGRETE GUTIÉRREZ, M. C.; CALVA GONZÁLEZ, J. J. (Comp.). **Desarrollo de colecciones y diseño de servicios**. México, DF: UNAM/Centro Universitario de Investigaciones Bibliotecológicas, 1996. p. 1-4.

OLIVEIRA, E. Identificando necessidades de acervo de usuários de um departamento acadêmico: uma abordagem bibliométrica. **Ciência da Informação**, Brasília, DF, v. 33, n. 1, p. 126-131, jan./abril 2004.

SMITH, L. C. Citation analysis. **Library Trends**, Illinois, v. 30, n. 1, p. 83-106, Summer 1981.

SPINAK, E. Indicadores cientificos. **Ciência da Informação**, Brasília, DF, v. 27, n. 2, p. 141-148, maio/ago. 1998. Disponível em: <http://www.scielo.br/pdf/ci/v27n2/spinak.pdf>. Acesso em: 07 jan. 2007.

SILBERSCHATZ, A., KORTH, H. F., S. SUDARSHAN, S. **Sistema de Bancos de Dados**. 3. ed. São Paulo: Makron Books, 1999. 778 p.

SILBERSCHATZ, A., KORTH, H. F., S. SUDARSHAN, S. **Sistema de Bancos de Dados**. 5. ed. São Paulo: Makron Books, 2005. 784 p.

SYLVIA, M. Citation analysis as an unobtrusive method for journal collection evaluation using psychology student research bibliographies. **Collection Building: studies in the development and effective use of library resources**, New York, v. 17, n. 1, p. 20-28, Mar. 1998.

TARAPANOFF, K. (Org.). **Inteligência organizacional e competitiva**. Brasília, DF: Editora da UnB, 2001.

TARAPANOFF, K.; MIRANDA, D. M.; ARAÚJO JÚNIOR, R. H. de. (Colab.). **Técnicas para tomada de decisão nos sistemas de informação**. Brasília, DF: Thesaurus, 1995.

TUÑÓN, J.; BRYDGES, B. Improving the quality of university libraries through citation mining and analysis using two new dissertation bibliometric assessment tools. *In: WORD LIBRARY AND INFORMATION CONGRESS. IFLA GENERAL CONFERENCE AND COUNCIL*, 71., 2005, Oslo. **Libraries: a voyage of discovery**. Disponível em: http://www.ifla.org.sg/IV/ifla71/papers/078e-Tunon_Brydges.pdf. Acesso em: 07 jan. 2007.

URBANO SALIDO, C. Tipología documental citada en tesis doctorales de informática: bases empíricas para la gestión equilibrada de colecciones. **Biblioteconomia i Documentación**, Barcelona, n. 5, dic. 2000. Disponível em: <http://www.ub.es/bid/05urban2.htm>. Acesso em: 17 jan. 2007.

_____. El análisis de citas en trabajos de investigadores como método para el estudio del uso de información en bibliotecas. **Anales de Documentación**, Murcia, v. 4, p. 243-266, 2001. Disponível em: <http://www.um.es/fccd/anales/ad04/ad0400.html>. Acesso em 17 jan. 2007.

VANZ, S. A. S.; CAREGNATO, S. E. Estudos de citação: uma ferramenta para entender a comunicação científica. **Em Questão**, Porto Alegre, RS, v. 9, n. 2, p. 295-307, jul./dez. 2003.

VERGUEIRO, W. C. S. Desenvolvimento de coleções: uma nova visão para o planejamento de recursos informacionais. **Ciência da Informação**, Brasília, DF, v. 22, n. 1, p. 13-21, jan./abr. 1993.

WILLIAMS, V. K.; FLETCHER, C. L. Materials used by master's students in engineering and implications for collection development: a citation analysis. **Issues in Science and Technology Librarianship**, Chicago, n. 45, Winter 2006, Disponível em: <http://www.istl.org/06-winter/refereed1.htm>. Acesso em: 07 jan. 2007.

RESUMO ESTENDIDO

SALES, André Luiz. **Análise de referências em teses como apoio à tomada de decisão no desenvolvimento de coleções da biblioteca universitária**. 2007. 48 p. Monografia (Graduação em Ciência da Computação) – Universidade Federal de Lavras, Lavras, MG⁸.

O trabalho aqui apresentado realiza o estudo do uso da informação técnica e científica, por meio da análise das citações utilizadas na elaboração de teses acadêmicas. A pesquisa de cunho quantitativo vale-se da bibliometria para apoiar a tomada de decisão no desenvolvimento de coleções de uma biblioteca. Analisa-se o conjunto de dados extraídos das teses defendidas entre os anos de 1999 a 2006, orientadas por professores pertencentes ao Departamento de Ciências Administrativas da Escola de Administração da Universidade Federal do Rio Grande do Sul (UFRGS) representando, até esta data, um total de 80 teses disponibilizadas em texto completo no catálogo *on-line* (*Online Public Access Catalogue - OPAC*) do Sistema de Bibliotecas da UFRGS. O objetivo geral é criar um banco de dados das referências bibliográficas de teses de um programa de pós-graduação e posteriormente estudar o uso da informação técnica e científica por meio da análise das citações. Os objetivos específicos coletar as teses disponíveis em formato digital, que irão servir de base para a análise; modelar e desenvolver um banco de dados das referências bibliográficas utilizadas em todas as teses; e criar consultas no banco de dados permitindo analisar a tipologia dos materiais bibliográficos utilizados nas citações, sua obsolescência, os títulos de periódicos citados e o idioma dos materiais consultados. Acredita-se que as informações obtidas da análise contribuem para melhorar a distribuição dos recursos financeiros disponíveis para investimento em material informacional, traçar diretrizes para o armazenamento retrospectivo da coleção, determinar a composição da coleção-núcleo de periódicos, avaliar em que idiomas devem ser selecionados os materiais bibliográficos para futuras aquisições por compra, permuta ou doação, bem como para preparar possíveis cortes na coleção devido a eventuais reduções no orçamento da biblioteca.

Palavras-chave: Banco de Dados. Análise de citações. Desenvolvimento de coleções.

⁸**Orientadora:** Profa. Olinda Nogueira Paes Cardoso - UFLA