

**IGOR CAMPOS LEAL**

**Análise de Citações da Produção Científica de uma Comunidade:  
A Construção de uma Ferramenta e sua Aplicação em um  
Acervo de Teses e Dissertações do PPGCI-UFMG**

**Belo Horizonte**

**2005**

**IGOR CAMPOS LEAL**

**Análise de Citações da Produção Científica de uma Comunidade:  
A Construção de uma Ferramenta e sua Aplicação em um  
Acervo de Teses e Dissertações do PPGCI-UFMG**

**Dissertação apresentada ao Curso de  
Mestrado da Escola de Ciência da Informação  
da Universidade Federal de Minas Gerais,  
como requisito parcial à obtenção do título de  
Mestre em Ciência da Informação.**

**Linha de Pesquisa: Organização e Uso da  
Informação  
Orientador: Prof. Dr. Marcello Peixoto Bax  
Co-orientadora: Profa. Dra. Lídia Alvarenga**

**Belo Horizonte**

**Programa de Pós-Graduação em Ciência da Informação  
Escola de Ciência da Informação  
Universidade Federal de Minas Gerais**

**2005**

Leal, Igor Campos

I24a      Análise de citações da produção científica de uma comunidade [manuscrito] : a construção de uma ferramenta e sua aplicação em um acervo de teses e dissertações do PPGCI-UFMG / Igor Campos Leal. - 2005.  
94 f.: il.

Orientador: Marcello Peixoto Bax.

Co-orientadora: Lídia Alvarenga.

Dissertação (mestrado) – Universidade federal de Minas Gerais, Escola de Ciência da Informação.

Bibliografia: f. 76-79.

Anexos: f. 80-94.

1. Bibliometria - Citações - Teses 2. Ciência da informação - Teses 3. Bibliotecas digitais – Teses I. Título II. Bax, Marcello Peixoto III. Alvarenga, Lídia IV. Universidade Federal de Minas Gerais. Escola de Ciência da Informação.

CDU: 02:004

Ficha Catalográfica: Biblioteca Etelvina Lima, Escola de Ciências da Informação da UFMG



**UFMG**

**Universidade Federal de Minas Gerais  
Escola de Ciência da Informação  
Programa de Pós-Graduação em Ciência da Informação**

**FOLHA DE APROVAÇÃO**

***"ANÁLISE DE CITAÇÕES DA PRODUÇÃO CIENTÍFICA DE UMA COMUNIDADE: A CONSTRUÇÃO DE UMA FERRAMENTA E SUA APLICAÇÃO EM UM ACERVO DE TESES E DISSERTAÇÕES DO PPGCI-UFMG"***

Igor Campos Leal

Dissertação submetida à Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Minas Gerais, como parte dos requisitos à obtenção do título de "Mestre em Ciência da Informação", linha de pesquisa "Organização e Utilização da Informação (OUI)".

Dissertação aprovada em: 15 de julho de 2005.

Por:

Prof. Dr. Marcelo Peixoto Bax – Orientador

Profa. Dra. Lídia Alvarenga – ECI/UFMG

Prof. Dr. Osvaldo Sérgio Farhat de Carvalho – DCC/UFMG

Aprovada pelo Colegiado do PPGCI

Profa. Maria Eugênia Albino Andrade

Coordenadora

Versão final Aprovada por

Prof. Marcelo Peixoto Bax

Orientador

## AGRADECIMENTOS

Pensar em pessoas a agradecer nesse momento não é algo fácil, pois, com certeza, tantos foram aqueles que participaram, colaboraram, me compreenderam e me apoiaram durante os 2 anos e meio desse projeto de pesquisa.

O primeiro não poderia ser outro que não o meu Deus, querido Pai tão presente em todos os momentos, não me deixando desanimar, mesmo com minha humana vontade de desistir. Esta é a grande certeza da minha vida: “Tudo posso Naquele que me fortalece”, Fil 4-13.

Dri, meu amor, amada noiva, futura esposa, devo agradecê-la desde o primeiro dia em que estivemos juntos. Se hoje sou quem eu sou, fiz o que eu fiz, cheguei aonde eu cheguei, o mérito não é apenas meu, pois você é parte integrante de todos os momentos, sem exceção. Obrigado por fazer parte da minha vida integralmente.

Meus amados papai e mamãe, uma pena não tê-los mais próximos, mas o amor e as orações de vocês, com certeza, me ajudaram a alcançar meus objetivos. Minha vida presente e futura são frutos colhidos do plantio do passado, com todo o carinho, amor, dedicação e apoio dados a mim durante toda a minha vida.

Queridos irmãos, Keila e Wallace, juntamente com Márcio e a minha linda Júlia. Vocês fazem parte da minha história de vida e obrigado por celebrarem esta vitória junto comigo.

Luzia e Eduardo, o apoio e o carinho de vocês foram essenciais à conclusão deste trabalho e a todos os outros projetos da minha vida que tenho traçado.

Aos meus eternos amigos, Pedro, Henrique e Thais. Sem vocês na minha vida, as coisas não seriam tão boas como têm sido. Vocês formam o meu porto seguro.

Meus amigos construídos com suor e esforço durante a minha vida belohorizontina, sob a ajuda de Deus e de Nossa Senhora Mãe da Igreja, obrigado por tudo! Rodrigão, Clá, Paulinha Costa, Rafinha, Lets, Lelé, Paulinha Santiago, e tantos outros, vocês são muito especiais...

Meus amigos de trabalho que tanto colaboraram com os seus conhecimentos técnicos e compreenderam minha ausência durante este período, muito obrigado.

Devo muito agradecer meus prezados tutores, Professor Marcello Peixoto Bax, pelos 5 anos de convivência e orientação e Professora Lídia Alvarenga, por me fazer acreditar que meus sonhos podem se tornar realidade.

Ao corpo técnico da Escola, principalmente a Viviany e a Goreth, por sempre estarem à disposição de forma muito prestativa.

E a todos os outros que fizeram parte desta história. Muito obrigado!

## RESUMO

Com a atual evolução tecnológica, a Internet e as novas formas de comunicação, estamos presenciando uma explosão informacional; constatação sem novidades, principalmente na área científica. Os pesquisadores nunca tiveram tanta facilidade no acesso ao material científico produzido. Porém, mecanismos de análise de todo esse material não são ainda tão acessíveis. Dessa forma, este trabalho apresenta o desenvolvimento de uma ferramenta de análise de citações, área de estudo da disciplina Tratamento da Informação conhecida como Bibliometria. O trabalho foi dividido em duas fases: a primeira diz respeito à implementação da ferramenta e a segunda à parte experimental. Trabalhou-se com um universo de 55 teses e dissertações defendidas e aprovadas no Programa de Pós-Graduação em Ciência da Informação da UFMG, com o objetivo de verificar a aplicabilidade da ferramenta desenvolvida. Com esses documentos foi construída uma biblioteca digital, utilizando a plataforma de criação de bibliotecas digitais Dspace, desenvolvido pelo MIT e HP, como base. Assim, um novo módulo foi acrescentado ao Dspace, o de análise de citações, em que todas as 4.495 referências bibliográficas presentes nos 55 documentos foram analisadas pelos indicadores: ano de publicação, tipo de publicação, país de origem, autor do item referenciado, título e idioma utilizado. Tais indicadores puderam ser analisados subdividindo-os por linha de pesquisa, orientador, geral e para um único documento. A partir dos gráficos gerados pela ferramenta, percebe-se o poder da análise bibliométrica. Conclui-se que, mecanismos como o apresentado, são necessários para o levantamento de insumos para análise da comunicação científica e estudos epistemológicos sobre os campos específicos do conhecimento no Brasil. Cabe ressaltar que a ferramenta seria potencializada se utilizada por outras comunidades científicas, de modo a apresentar a toda comunidade científica de uma determinada área um ponto único de análise do que é produzido nos níveis de pós-graduação *stricto sensu* no Brasil.

Palavras-chave: citação, bibliometria, análise de citações, PPGCI, UFMG, Dspace, bibliotecas digitais.

## ABSTRACT

Due to the current technological evolution, internet facilities and such wide range of media communication, an “information explosion” (or “information bug”) has been growing up lately, which is not a newness, specially in scientific area. Researchers have never had so much facility accessing the published scientific material. However, analysis mechanisms of the mentioned material are still not that accessible. Based on these points this work will present the development of a citation analysis tool, area of study of the discipline Information Processing, better known as bibliometrics. It was divided in two phases: the first one refers to the tool implementation and the second one to the experimental part. It has been worked on the universe of 55 thesis and dissertations approved by the UFMG Information Science Post-Graduate Program, with the purpose to verify the usability of the developed tool. Trough these documents was created a digital library, using as its base the framework of digital library creation named Dspace, which was developed by MIT and HP. A new module was added to Dspace, the citation analysis, trough wich all the 4.495 bibliographic references present in the 55 documents were analyzed according to the following indicators: published year, publication type, origin country, author of the cited work, title and language used. Such indicators could be analyzed subdividing by subarea of research, advisor, general and for a single document. The charts generated by the developed tool have demonstrated the power of the bibliometrics analysis. The dissertation concludes that techniques like this are important and necessary to the survey of the bases for analyze the scientific communication and studies about the epistemology of specific knowledge areas in Brazil. This tool would be potencialized if it used by others scientific communities, in order to give them a unique place of analysis of what they have produced in the *stricto sensu* postgraduate levels in Brazil.

Key-words: citation, bibliometrics, citation analysis, PPGCI, UFMG, Dspace, digital libraries.

## LISTA DE GRÁFICOS

Gráfico 1 – Crescimento da literatura científica.....	11
Gráfico 2 – Análise Geral das citações por Ano .....	49
Gráfico 3 – Análise Geral das citações por Idioma .....	49
Gráfico 4 – Análise Geral das citações por País de Publicação .....	50
Gráfico 5 – Análise Geral das citações por Tipo de Publicação .....	50
Gráfico 6 – Análise Geral das citações por Autor .....	50
Gráfico 7 – Análise Geral das citações por Assunto dos Títulos .....	51
Gráfico 8 – Análise das citações por Ano para a Linha de Pesquisa GIC.....	52
Gráfico 9 – Análise das citações por Idioma para a Linha de Pesquisa GIC .....	52
Gráfico 10 – Análise das citações por País de Publicação para a Linha de Pesquisa GIC.....	52
Gráfico 11 – Análise das citações por Tipo de Publicação para a Linha de Pesquisa GIC.....	53
Gráfico 12 – Análise das citações por Autor para a Linha de Pesquisa GIC .....	53
Gráfico 13 – Análise das citações por Assunto dos Títulos para a Linha de Pesquisa GIC.....	53
Gráfico 14 – Análise das citações por Assunto das Referências para a Linha de Pesquisa GIC ..	54
Gráfico 15 – Análise das citações por Ano para a Linha de Pesquisa ICS .....	54
Gráfico 16 – Análise das citações por Idioma para a Linha de Pesquisa ICS .....	55
Gráfico 17 – Análise das citações por País de Publicação para a Linha de Pesquisa ICS .....	55
Gráfico 18 – Análise das citações por Tipo de Publicação para a Linha de Pesquisa ICS .....	55
Gráfico 19 – Análise das citações por Autor para a Linha de Pesquisa ICS .....	56
Gráfico 20 – Análise das citações por Assunto dos Títulos para a Linha de Pesquisa ICS .....	56
Gráfico 21 – Análise das citações por Assunto das referências para a Linha de Pesquisa ICS ....	56
Gráfico 22 – Análise das citações por Ano para a Linha de Pesquisa OTI.....	57
Gráfico 23 – Análise das citações por Idioma para a Linha de Pesquisa OTI .....	57
Gráfico 24 – Análise das citações por País de Publicação para a Linha de Pesquisa OTI.....	58
Gráfico 25 – Análise das citações por Tipo de Publicação para a Linha de Pesquisa OTI.....	58
Gráfico 26 – Análise das citações por Autor para a Linha de Pesquisa OTI .....	58
Gráfico 27 – Análise das citações por Assunto dos Títulos para a Linha de Pesquisa OTI.....	59
Gráfico 28 – Análise das citações por Assunto das Referências para a Linha de Pesquisa OTI ..	59
Gráfico 29 – Análise das citações por Ano para o Orientador selecionado .....	60
Gráfico 30 – Análise das citações por Idioma para o Orientador selecionado.....	60
Gráfico 31 – Análise das citações por País de Publicação para o Orientador selecionado .....	60
Gráfico 32 – Análise das citações por Tipo de Publicação para o Orientador selecionado .....	61
Gráfico 33 – Análise das citações por Autor para o Orientador selecionado.....	61
Gráfico 34 – Análise das citações por Assunto dos Títulos para o Orientador selecionado .....	61
Gráfico 35 – Análise das citações por Assunto das Referências para o Orientador selecionado..	62



## LISTA DE TABELAS

Tabela 1 – Razões para a citação de Weinstock (1971) .....	25
Tabela 2 – Áreas de aplicação da Bibliometria (SENGUPTA, 1992).....	28
Tabela 3 – Variáveis independentes .....	39
Tabela 4 – Mapeamento: informações do documento x metadado Dublin Core .....	39
Tabela 5 – Variáveis dependentes .....	40
Tabela 6 – Universo de trabalho.....	47

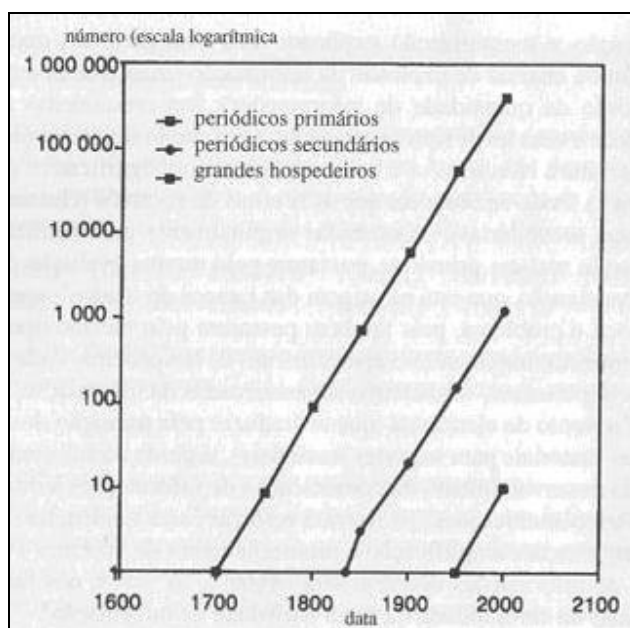
## SUMÁRIO

1	Introdução.....	11
1.1	Problemas .....	13
1.2	Objetivos.....	14
1.3	Justificativa.....	15
2	Fundamentação Teórica.....	17
2.1	Bibliotecas Digitais como instrumentos de apoio .....	17
2.2	Comunicação Científica e sua necessidade para o avanço da Ciência .....	20
2.3	Citação como um fato social no contexto do sujeito .....	23
2.4	Análise de citação: o que podemos descobrir?.....	26
2.5	Índices de citação: uma ferramenta de monitoração da produção científica.....	30
2.5.1	Institute for Scientific Information (ISI) .....	32
2.5.2	Eugene Garfield.....	34
2.6	Utilização de indicadores científicos no apoio à tomada de decisão.....	35
3	Metodologia.....	37
3.1	Etapa de implementação.....	37
3.1.1	Variáveis independentes .....	38
3.1.2	Variáveis dependentes .....	40
3.1.3	Fases da implementação .....	41
3.2	Etapa experimental .....	44
4	Apresentação dos resultados.....	48
4.1	Resultados Gerais .....	48
4.2	Resultados por Linha de Pesquisa .....	51
4.2.1	Gestão da Informação e do Conhecimento - GIC.....	51
4.2.2	Informação, Cultura e Sociedade - ICS .....	54
4.2.3	Organização e Tratamento da Informação - OTI.....	57
4.3	Resultados por Orientador .....	59
5	Análise dos resultados .....	63
5.1	Análise Geral .....	64
5.2	Análise por Linha de Pesquisa.....	66
5.3	Análise por Orientador .....	68
6	Considerações Finais .....	70
7	Trabalhos Futuros .....	73
8	Referências Bibliográficas.....	76
	ANEXOS.....	80

## 1 Introdução

Nos anos que precederam à Internet, a comunidade científica e tecnológica ressentia-se da dificuldade de acesso à literatura científica. Primeiro pelo fato do acesso ser muito dispendioso, dado ao elevado custo das bases de dados científicas mais importantes, em seguida, pelo fato das tecnologias de recuperação de informação e comunicação serem ainda rudimentares, naquela época, não sendo possível o acesso direto a tais informações.

Com o avanço das tecnologias da informação e da comunicação e o advento da Internet, o registro e a disseminação da informação científica e tecnológica sofreram grandes impactos como, por exemplo, o crescimento exponencial do volume de informações registradas, como visto no gráfico abaixo.



**Gráfico 1 – Crescimento da literatura científica**

Fonte: A Ciência da informação, Yves-François Le Coadic, p. 8

Outro impacto relevante foi o fato de que o usuário passou a ter acesso direto à informação, sem a intermediação de um profissional da informação, inclusive para os autores, que passaram a ter maior facilidade em tornar disponíveis os resultados de suas pesquisas, livrando-se, de certa

maneira, da dependência dos grandes editores de revistas e jornais científicos. Nesse contexto, para encontrar informações relevantes, é necessária a utilização de técnicas, ferramentas e instrumentos que auxiliem a recuperação de informação.

Hoje várias soluções para esses problemas têm sido aplicadas e, neste trabalho, apresentamos uma ferramenta que objetiva colaborar com a divulgação e disseminação dos produtores do conhecimento científico nacional - o índice de citação.

Toda citação encontrada em um documento científico é feita para itens anteriores, tendo o principal intuito de fornecer sustentação às teorias a serem trabalhadas (CRONIN, 1984). Essas citações são compostas de dados que, sob análise cuidadosa, podem se tornar informações preciosas. Citações são compostas por aspectos temporais, espaciais, autorais, editoriais etc. e a análise conjunta ou separada de cada um desses aspectos pode gerar informações sobre a essência do material científico produzido.

Várias funcionalidades podem ser construídas a partir das citações de um documento científico. Talvez a principal delas tenha sido definida por Lancaster nos termos “encontre, para um determinado item bibliográfico que seja do conhecimento de quem faz a busca, itens posteriores que o tenham citado” (1993, p. 166). Além dessa funcionalidade, outras podem interessar, como “a avaliação de artigos, de autores, etc., e a análise de tendências da pesquisa científica” (LAWRENCE; GILES; BOLLACKER, 1999, p. 1). Com essa análise, extraem-se informações relevantes que conduzem a conclusões que ajudam a inferir sobre a situação atual de uma comunidade científica; como um “retrato” da produção científica da comunidade. Esse “retrato” permite conhecer detalhes da produção dos membros dessa comunidade e seu acompanhamento sistemático viabiliza o autoconhecimento. Medidas podem, então, ser tomadas a fim de se reforçar ou corrigir determinados aspectos, com melhoria contínua da qualidade científica da comunidade.

Para um programa de pós-graduação, buscar o autoconhecimento é o caminho necessário para se alcançar a excelência em sua área de atuação; tanto para a satisfação do corpo discente, por estar sendo formado em uma instituição preocupada com sua evolução, quanto para o corpo docente, por estar atuando de forma pró-ativa nessa formação.

A análise das referências permite descobrir quais têm sido as origens do conhecimento produzido, quais as bases e os cerne utilizados na construção do conhecimento produzido pelos

futuros mestres e doutores. De posse de ferramentas tecnológicas adequadas, os “retratos” tirados podem ser atualizados a cada novo documento científico adicionado.

## 1.1 Problemas

Todo pesquisador convive com a difícil tarefa de encontrar o que realmente irá satisfazer suas necessidades informacionais em meio a todo conhecimento já produzido. Nessa condição falta a este usuário ferramentas que correspondam melhor ao estado tecnológico atual, fazendo uso efetivo das novas Tecnologias da Informação e Comunicação (TIC's).

Apesar da existência de meios para se obter tais conhecimentos, já que todo pesquisador conectado à Internet tem acesso a várias fontes, como bibliotecas, CD-Roms, bases de dados on-line, *sites* na Internet, periódicos eletrônicos, entre vários outros, o grande problema consiste na falta de intercâmbio entre esses meios de comunicação, como a falta de mecanismos de ligação entre os *sites* de periódicos nacionais.

Há de se atentar para as iniciativas nacionais nesse intuito. Podemos citar o projeto Scielo (PACKER, 1998), o *Scientific Electronic Library Online*, uma iniciativa nacional para promover o acesso livre a documentos eletrônicos. Tal projeto tem hoje registrado cerca de 298 periódicos de acesso gratuito na internet, publicados no Brasil, Chile, Cuba, Venezuela, México, Costa Rica e Espanha, cobrindo diversas áreas, das Ciências Biológicas às Artes.

Outro exemplo é o Programa Prossiga (PROSSIGA, 2005). Criado em 1995, o programa tem por objetivo promover a criação e o uso de serviços de informação na Internet voltados para as áreas prioritárias do Ministério da Ciência e Tecnologia, assim como estimular o uso de veículos eletrônicos de comunicação pelas comunidades dessas áreas.

Outro projeto de significância é o Portal de Periódicos da Capes, acessado pelo endereço <http://www.periodicos.capes.gov.br>. As comunidades de 152 instituições brasileiras de ensino superior e de pesquisa têm acesso à produção científica mundial atualizada, com textos completos de artigos de mais de 9.000 revistas nacionais e internacionais e mais de 90 bases de dados referenciais com resumos de documentos em todas as áreas do conhecimento.

Em todos os exemplos citados, o usuário é capaz de trabalhar com documentos provenientes de várias fontes distintas através de apenas um ponto de entrada. Mas, mesmo nesses casos, ainda se

carece de sistemas de recuperação da informação que englobem “toda” a produção científica de uma determinada comunidade.

Apesar de tais esforços, durante esta pesquisa não foram encontradas ferramentas nacionais que trabalhassem com conteúdo científico nacional disponibilizadas na internet com o intuito de agregar valor analítico. Grande parte dos *sites* das maiores universidades brasileiras já possui mecanismos de recuperação dos documentos produzidos em seus programas de pós-graduação, até mesmo em texto completo, mas, infelizmente, o usuário tem em mãos pontos de acessos, sem possibilidades de maiores análises sobre os documentos.

Os exemplos de instituições internacionais como o *Institute for Scientific Information* (ISI) e o CiteSeer, que têm trabalhado neste mesmo intuito, têm mostrado algo que já era de se esperar e que não é mais novidade. Sistemas internacionais normalmente privilegiam a produção técnico-científica dos países desenvolvidos. O ISI, por exemplo, contemplava apenas 0,37% dos periódicos nacionais, em 1999 (TARGINO; GARCIA, 2000). Dessa forma, o autoconhecimento de uma comunidade científica brasileira nunca seria obtido a partir do uso dessas ferramentas. Essa realidade abre uma porta para vários trabalhos, como esta dissertação, que tenham como premissa a idéia de privilegiar o conhecimento que serve de base ao que é produzido em comunidades científicas, assim como seu impacto ambos por meio de análise das citações.

## **1.2 Objetivos**

Este trabalho tem dois objetivos principais. O primeiro é o de criar uma biblioteca digital dotada de uma ferramenta de análise de citações de teses e dissertações. Para se tornar possíveis os estudos bibliométricos das citações presentes, foram levantados indicadores da produção científica dos documentos constituintes desta biblioteca digital, com o objetivo de que, com estes indicadores, torne possível o autoconhecimento sobre a produção de uma comunidade específica e que permita um conhecimento mais profundo e analítico dessa produção científica.

O segundo objetivo, alcançado a partir do primeiro, se refere ao efetivo uso da ferramenta para, a partir da biblioteca digital criada, fazer um desenho da situação da produção do Programa de Pós-Graduação em Ciência da Informação (PPGCI) da Escola de Ciência da Informação (ECI) da

Universidade Federal de Minas Gerais (UFMG), efetuando análises a fim de obter respostas a algumas questões, como, por exemplo:

1. Quais são os autores mais citados pelas teses e dissertações defendidas no PPGCI?
2. Como se dá a distribuição das citações nesses mesmos documentos durante os anos?
3. Quais são os assuntos de pesquisa mais freqüentes nessas teses e dissertações?
4. Todas essas questões se comportam de maneiras semelhantes nas três linhas de pesquisa do PPGCI?
5. Como é o comportamento dessas questões se analisarmos documentos separadamente por orientador?

O objetivo da ferramenta é permitir à comunidade do PPGCI um acompanhamento contínuo dessas questões, com o uso e atualização constante do estudo analítico do acervo que compõe sua produção bibliográfica de teses e dissertações.

### **1.3 Justificativa**

A pesquisa em questão justifica-se pela necessidade de ferramentas mais eficazes que contemplem o material produzido por uma comunidade e que acrescente aos seus usuários análises mais contundentes de como tem sido a sua produção científica, uma vez que há um crescimento contínuo do volume da informação técnico-científica nacional.

A ferramenta aqui explicitada possui considerável potencial de análise, podendo aprofundar, ainda mais, nas perguntas e respostas que ela pode oferecer, relacionando outros indicadores, fazendo cruzamentos entre os indicadores, o que for necessário e conveniente para a comunidade em que esta atuará. Além disso, ela foi construída de forma que seu uso não se limite apenas a trabalhos da área da Ciência da Informação, pois esse universo foi selecionado como parte da Metodologia do trabalho. Isso possibilita a transferência deste conhecimento para qualquer comunidade científica que se dispuser a realizar as etapas necessárias às análises.

Além dessas, pode-se incluir a obrigação de se privilegiar a produção científica nacional. A comunidade brasileira carece de sistemas que façam tal abordagem, além de ser de suma importância o levantamento de indicadores da produção científica de uma comunidade como

norteador para as tomadas de decisão, já que, com a análise das citações, podem ser feitas descobertas interessantes.

Por fim, vale ressaltar que esta pesquisa é a finalização de uma parte de um projeto financiado pelo CNPq, denominado SABiO<sup>1</sup> (Sistema de Agentes para Bibliotecas Online) (BAX, 2003).

---

<sup>1</sup> Processo CNPq no. 403170/03-6.



## **2 Fundamentação Teórica**

### **2.1 Bibliotecas Digitais como instrumentos de apoio**

A fim de minimizar os problemas de explosão informacional e da necessidade de se disponibilizar o conhecimento técnico-científico produzido, a comunidade científica tem trabalhado com o conceito de bibliotecas digitais, conforme pode ser visto nos trabalhos de Bush (1945), Buckland (1991) e Cardoso (1994).

O avanço das novas tecnologias na área de bibliotecas e nas TIC's é evidente. Esse momento torna viável a utilização desses “novos” meios tecnológicos na tentativa de auxiliar a comunidade científica na busca de métodos melhores e mais eficientes de se trabalhar a seleção, aquisição, organização, armazenagem, recuperação, disseminação e futura avaliação de sua produção científica, permeando todas as funções básicas de um sistema de informação, conforme tratado na disciplina Tratamento da Informação. Dessa forma, as bibliotecas digitais se impõem como um fenômeno que pode vir a amenizar alguns dos problemas enfrentados pelos que pretendem resolver suas necessidades de informação por meio do contexto digital (DIAS, 2001).

Em busca de uma conceituação para bibliotecas digitais, Dias afirma que a

[...] a biblioteca digital está se firmando como a expressão que significaria, no contexto digital, um conjunto de artefatos, conhecimentos, práticas e uma comunidade, que engendra compromissos realísticos assumidos por profissionais da informação, analistas de sistemas e usuários (DIAS, 2001, p. 3).

Outro importante aspecto do conceito de bibliotecas digitais, segundo Alvarenga, se refere a um

[...] conjunto intencional de objetos concebidos em meio digital ou desmaterializados de sua condição física tradicional, constituídos de funções inteiramente novas que lhes garantem a hipertextualidade e caráter multimidiático, tornando-os passíveis de acatar novos e peculiares arranjos e tipos de abordagens no processo de sua recuperação, [ou também uma biblioteca digital é] um conjunto de objetos digitais construídos a partir do uso de instrumentos eletrônicos, concebidos com o objetivo de registrar e comunicar pensamentos, idéias, imagens e sons, disponíveis a um contingente ilimitado de pessoas, dispersas onde quer que a plataforma WWW alcance (ALVARENGA, 2001).

A desterritorialização dos documentos, que consiste no fato de o documento passar a ter sua materialidade desvinculada da forma física tradicional, o papel, é uma das principais mudanças ocorridas. Este fato corrobora a idéia de se trabalhar com bibliotecas digitais na Internet, principalmente quando estes documentos são os resultados dos trabalhos técnico-científicos de uma comunidade científica.

A possibilidade de disponibilizar tais trabalhos na rede Internet os torna mais citados, como já demonstrado por Lawrence (2001). Essa disponibilização beneficia os pesquisadores acadêmicos, conforme salienta Pacheco, pois oferece aos pesquisadores “ubiquidade, baixo custo e facilidades de pesquisa” (PACHECO; KERN, 2001, p. 65), fazendo com que sua produtividade tenda a aumentar.

Outro aspecto interessante discutido por Pacheco e Kern (2001) é o fato das universidades, notoriamente as públicas, terem o dever de dar um retorno à sociedade, uma vez que são os recursos públicos suas grandes fontes de financiamento. O autor trabalha principalmente com as teses e dissertações e, ao divulgar estes trabalhos defendidos pelos pós-graduandos, a sociedade tem um resultado concreto desse investimento.

A partir da possibilidade da sociedade ter acesso ao conteúdo das teses e dissertações através de bibliotecas digitais bem aparelhadas, a pós-graduação brasileira tem a oportunidade de expor a qualidade do que produz e, assim, promover sua integração com o restante da sociedade. Dessa forma, reforça-se o caráter de transparência e acesso público ao acervo nacional de conhecimento.

Podemos tomar esta área de estudo, o uso do meio eletrônico como suporte básico de uma biblioteca, como uma área de pesquisa que vem sendo bastante trabalhada. Soergel, relatou em um de seus trabalhos os principais temas de pesquisa que devem ser realizados na área de bibliotecas digitais (SOERGEL, 2002). Soergel aponta que é necessária uma “plataforma para ser utilizada como uma perspectiva das pesquisas existentes e como uma visão estruturada para o desenvolvimento de novas idéias” (SOERGEL, 2002, p. 1). Segundo o autor, há três princípios básicos a serem utilizados como guia em toda agenda de pesquisas em bibliotecas digitais. No primeiro Soergel afirma que devem prover praticidade; no segundo elas devem propiciar novos meios para o trabalho intelectual e, por fim, devem dar suporte também à colaboração e às comunidades na prática.

A fim de se trabalhar com todos os aspectos discutidos por Soergel, deve-se procurar soluções que auxiliem e, assim, evitem a difícil tarefa de descobrir, fazer o que já foi descoberto e feito. Conforme Soergel também afirma, deve-se buscar outros centros de pesquisa que já tenham avançado em algum ponto específico e utilizar seus produtos. Na era do software livre, não há mais como fugir dos sistemas de código aberto, no estilo do sistema operacional Linux. Soergel (2002) também apresenta uma lista de centros de pesquisa que vêm atuando dessa forma a fim de facilitar o trabalho inicial.

Seguindo estes pensamentos, dois grandes centros de pesquisa, um de uma empresa privada e outro de uma universidade, se uniram com a finalidade de entregar à comunidade, principalmente às universidades, uma ferramenta que auxiliasse na construção de bibliotecas digitais. A Hewlett & Packard (HP) e o *Massachusetts Institute of Technology* (MIT) trabalharam juntos na construção do Dspace<sup>2</sup>, um software livre sob a licença BSD<sup>3</sup>. Uma ferramenta implementada para funcionar totalmente na interface da Web, via Internet, com muitas funcionalidades complexas o suficiente para que justifique sua utilização (primeiro princípio de Soergel), pois fornece comunicação com banco de dados, controle de usuários, mecanismos de busca, fácil administração de coleções e de usuários, comunicação entre os usuários através de comunidades (terceiro princípio de Soergel), possibilidade dos usuários adicionarem documentos à coleção, entre outros.

A partir dessa ferramenta, todo processo de criação de uma coleção de objetos, após a sua instalação e configuração, se restringe a adicionar cada documento propriamente dito nos mais variados formatos (MS Word, PDF, PostScript, HTML, texto puro, por exemplo), criando pontos de acesso ao objeto através dos metadados definidos pelo padrão Dublin Core (DUBLIN CORE, 2005).

Com as principais funcionalidades de um sistema de gerenciamento de uma biblioteca digital (como o Dspace) já implementadas, pode-se dar mais atenção às especificidades de cada instituição, liberando a equipe para trabalhar com o que pretende adicionar ao sistema. No caso do presente trabalho, é a criação do índice de citações de cada trabalho defendido junto ao Programa de Pós-Graduação em Ciência da Informação da Escola de Ciência da Informação da UFMG presente na biblioteca, para que o usuário possa obter não apenas a informação que

---

<sup>2</sup> Disponível em <http://www.dspace.org>. Acessado em 20/07/2004.

<sup>3</sup> Disponível em <http://www.opensource.org/licenses/bsd-license.php>. Acessado em 13/08/2004.

deseja, através de um documento recuperado como resposta a uma consulta, mas, também, uma análise bibliométrica de suas citações (alcançando o segundo princípio de Soergel).

## **2.2 Comunicação Científica e sua necessidade para o avanço da Ciência**

Segundo John Ziman, a comunicação é essencial para a pesquisa científica, pois a ciência é caracterizada como um “conhecimento público”, cujo principal objetivo deve ser “criar, criticar ou contribuir com um consenso racional de idéias e informações” (ZIMAN, 1969). Em outro texto, Ziman define a Ciência como “um produto consciente da humanidade, com suas origens históricas bem documentadas, um escopo e um conteúdo bem definidos”, além de “ser precisa, metódica, acadêmica, lógica e prática” (ZIMAN, 1979, p. 17).

Segundo o mesmo autor, toda ciência é cumulativa, derivada e publicada. Cumulativa é a propriedade da ciência de continuar sempre progredindo, utilizando o conhecimento anterior para a produção de outros novos. Derivada porque, exceto as grandes invenções, sempre partimos de algo, onde não é preciso o autor “reinventar a roda”. E a ciência é publicada pois, se não fosse assim, o resultado de uma pesquisa nunca seria assimilado pela comunidade, fazendo com que não ocorresse nenhuma das duas propriedades anteriores. Isso a torna como uma atividade intrinsecamente social, moldada e determinada pelas relações sociais entre os indivíduos.

Confirmando essas características, Ziman afirma que

Os fatos e teorias do conhecimento científico têm de passar por um crivo, por uma fase de análises críticas e de provas, realizadas por outros indivíduos competentes e desinteressados, os quais deverão determinar se eles são bastante convincentes para que possam ser universalmente aceitos. (ZIMAN, 1979, p. 24).

Sendo assim, podemos perceber que um conhecimento só é aceito pela comunidade a partir do momento em que algum tipo de avaliação se realizar. Após essa validação, esse conhecimento poderá se tornar público através das várias formas de comunicação existentes no meio científico.

Ao construir conhecimentos a partir de outros anteriores, pesquisadores fazem uso da característica de ser cumulativa. Para tanto, os pesquisadores devem fazer uso intensivo de referências e citações, pois é quase impossível escrever ou publicar trabalhos sem basear-se em outros anteriores relacionados. Segundo Ziman, “uma obra sem notas de rodapé ou uma

bibliografia torna-se imediatamente suspeita, sendo um indício certo de que se trata da obra de um excêntrico ou de um novato” (ZIMAN, 1979, p. 72).

Outro autor que relata sobre a relevância da comunicação para a comunidade científica e para o avanço da ciência é Meadows. Em um de seus trabalhos mais recentes sobre comunicação científica, ele inicia seu discurso com a afirmação que “a comunicação situa-se no próprio coração da ciência” (MEADOWS, 1999, p. 1). Neste estudo, Meadows afirma que a comunicação eficiente e eficaz é parte essencial do processo de investigação científica.

Essa ênfase dada à comunicação Meadows já estabelecia em trabalhos mais antigos, ao afirmar, em 1979, que a

[...] ciência e a comunicação parecem caminhar intrinsecamente juntas, tanto que a produção e a disseminação dos resultados de pesquisas estão lado a lado. Conseqüentemente, o crescimento científico tem envolvido não somente o aumento dos esforços humanos e os apoios financeiros, como também uma vasta proliferação de publicações (MEADOWS, 1979, p. 1).

Tomando, ainda, as idéias de Meadows, em temas de comunicação, as duas características mais importantes do pesquisador são a quantidade de informações que comunica e sua qualidade. Ele afirma que “uma medida de quantidade, no caso de pesquisadores acadêmicos, é o número de artigos de periódicos que publicam” (MEADOWS, 1999, p. 85). Sobre o outro aspecto, “uma forma de avaliar a qualidade consiste em avaliar o nível de interesse dos outros pela pesquisa. O método mais simples é por meio de citações dessa pesquisa na bibliografia subsequente” (MEADOWS, 1999, p. 89).

Pode-se aprofundar ainda mais nessa análise ao citar o filósofo Jurgen Habermas, quando ele escreve sobre a importância da comunicação para a evolução científica, sendo o progresso científico uma característica exemplar da ciência (HABERMAS, 1972).

Habermas também trata da importância da vivência anterior para a produção de um conhecimento novo em cada indivíduo:

A conexão da vida, expressão e entendimento [engloba] não somente os gestos, expressões e palavras com as quais cada homem se comunica ou permanentes criações espirituais ... ou a constante objetividade da mente nas formas sociais ...: mesmo a unidade filosófica da vida é conhecida por ela mesma através da relação da experiência e entendimento, tornando-se atento a ela mesma no presente e redescobrimo-se nas memórias de algo no passado... Em resumo, esse é o processo de entendimento através do qual a vida se torna iluminada sobre si mesma em sua profundidade. Por outro lado, nós entendemos a nós mesmos e aos outros somente por inserir nossa própria experiência de vida em todas expressões próprias e de outras vidas (HABERMAS, 1972, p. 148, grifo nosso).

Essa constatação de Jurgen Habermas nos dá pistas de que, somente através de manifestações comunicativas, a ciência permaneceu, permanece e permanecerá em evolução.

Ainda segundo Habermas, o diálogo estabelecido entre os especialistas técnicos e as instâncias de decisão pública deve considerar os interesses sociais e deve ser regido por seu quadro de valores éticos e morais. Para a instauração eficaz desses diálogos, uma circulação livre e ampla da comunicação científica se faz necessária e, para tanto, algumas barreiras ainda persistem e precisam ser removidas. A primeira delas diz respeito à passagem da informação científica ao grande público. Habermas sonha também com o resgate das premissas de universalização do conhecimento, da moral e da arte.

Meadows (1999) afirma que já no final do século XIX, havia uma crescente pressão sobre a pesquisa e a comunicação, levantando-se a questão de se tentar descobrir possibilidades para atender melhor as atividades de comunicação dos pesquisadores num ambiente tecnológico que passava por rápidas alterações. Ainda mais hoje, onde não apenas os membros do público em geral sentem dificuldades em ter acesso a informações científicas, mas é evidente a necessidade dos próprios pesquisadores de mais assistência para identificar, na massa de toda literatura disponível, o material de que necessitam. Urge a necessidade de identificarmos como interagir com a Tecnologia da Informação em constante desenvolvimento, de modo a tornar melhor o futuro da comunicação das pesquisas.

Meadows (1999) também destaca a importância da estruturação em rede para o estímulo do trabalho em equipe, uma vez que hoje não se é possível mais imaginar um pesquisador isolado do mundo, sem ter acesso à Internet. Dessa forma, segundo o autor, a possibilidade de todos terem

acesso aos mesmos dados e interagirem facilmente em sua utilização, favorece os esforços coletivos.

Neste contexto, posicionamos os índices de citação como um facilitador de toda a comunicação científica, levando a um progresso da mesma, conforme poderá ser visto nas seções seguintes.

### **2.3 Citação como um fato social no contexto do sujeito**

A citação pode ser entendida como um evento ou fato social, pois ela é feita levando-se em consideração todo o contexto do sujeito, com toda a sua experiência anterior, sua rede de conhecimento de outros autores, suas próprias indagações e as de outros.

Cronin afirma que “num mundo ideal, citações devem ser tidas como marcas ou símbolos objetivos, devem ser vistas como empregadas de modo consistente e devem ser suscetíveis a análises e interpretações consistentes” (1984, p. 68). Ele faz uma definição mais objetiva ao dizer que “citações bibliográficas têm sido descritas como símbolos, marcas, metáforas e sinais. Genericamente, citações atuam como sinais: elas denotam trabalhos particulares e implicam alguma forma de relacionamento entre o artigo citante e o citado” (CRONIN, 1984, p. 69). Cronin também afirma que as citações têm sido a forma mais comum de dar crédito e reconhecimento na Ciência e que o hábito de citar dos autores

mostra conformidade e consistência, uma vez que a citação é vista como uma atividade séria, governada por conhecimento tácito de como e porquê autores devem reconhecer o trabalho dos outros. As razões de um autor citar de um modo particular em um tempo particular são controladas por um conjunto de normas internalizadas. (CRONIN, 1984, p. 2).

Garfield, um dos maiores entusiastas da área de estudo de citações, define citação de uma forma objetiva, direta como “uma representação precisa, não ambígua de um assunto que não requer interpretação e é imune a mudanças terminológicas. Além disso, a citação retém sua precisão através do tempo, podendo ser usada em documentos escritos em línguas diferentes” (GARFIELD, 1979, p. 3). Já MacRoberts e MacRoberts assumem um conceito de citação mais voltado a um lado de mérito, homenagem, ao falarem da lista de referências bibliográficas como uma “lista de influência, onde autores citam para dar crédito onde o crédito é devido. As

referências citadas por um autor são fortes indicadores da influência em seu trabalho” (MACROBERTS e MACROBERTS, 1989, pág. 342).

Outra abordagem mais subjetiva é dada por Alvarenga (1996) ao afirmar que a citação é um procedimento de comunicação e transferência de informações entre cientistas. Confirmando esta abordagem social das citações, Alvarenga diz que:

Permeia o ato de citar, todo um espectro de implicações psicológicas, sociológicas, e históricas, assim como inclui influências de outras dimensões, tais como: o fenômeno de auto-citações (narcisismo); influências entre autores e instituições (presença do ‘já-dito’, influência, conhecimento baseado em conhecimentos previamente comunicados; adesão a paradigmas). (ALVARENGA, 1996, p. 83)

Seguindo este raciocínio, uma citação não ocorre por acaso e nem poderia ser, pois é através dela que o autor fundamenta a sua teoria podendo, para tanto, utilizar-se de várias justificativas na escolha de suas referências bibliográficas, entre elas podemos citar: a ação narcisista (auto-citação); a citação como discípulo, ao se buscar uma luz para a resolução de algum problema em outros autores; a citação como forma de elogiar; a citação para o mesmo grupo de pesquisa, criando os colégios invisíveis; e a criação de redes de pesquisadores, aqueles do mesmo assunto, que citam um ao outro para a evolução de suas pesquisas. Ziman (1979) afirma que delicadeza, cortesia, política e lealdade são possíveis justificativas para determinadas citações.

Por ser regida por “normas internalizadas”, regras próprias, que em certo momento anterior passaram a fazer parte do cotidiano do autor, o ato de citar pode ocorrer sob várias intenções e, às vezes, difíceis de se inferir. Vários foram os pesquisadores que se esforçaram no levantamento de motivações que levam um autor a selecionar trabalhos anteriores de determinados autores em detrimento de outros. Através da Tabela 1, vislumbra-se as razões enumeradas por Weinstock (1971).



1. Homenagear os pioneiros;
2. Dar crédito aos trabalhos relacionados;
3. Identificar metodologia, equipamento, etc.;
4. Prover formação;
5. Corrigir seu próprio trabalho;
6. Corrigir o trabalho de outros autores;
7. Criticar trabalhos anteriores;
8. Provar fatos;
9. Alertar pesquisadores de trabalhos futuros;
10. Prover orientação a trabalhos pouco disseminados, pouco indexados ou não citados;
11. Autenticar dados e classes de fatos – constantes físicas, etc.;
12. Identificar publicações originais em que idéias ou conceitos foram discutidos;
13. Identificar publicações originais que descrevam um conceito eponímico;
14. Negar trabalho ou idéias de outros; e
15. Disputar prioridade de outros.

**Tabela 1 – Razões para a citação de Weinstock (1971)**

Por essa gama de possibilidades, Cronin afirma que uma teoria das citações é difícil de ser construída, se não impossível, “pelo motivo da citação ser um ato privado e por haver muitas formas de interpretação em relação no que constitui uma citação correta ou aceitável” (CRONIN, 1984, p. 83). Em consequência disto, Cronin diz que a citação deve ser tomada como um processo, perpassando todo o conhecimento prévio do autor, inclusive o tácito, até a transformação dessa idéia mental em uma marca ou símbolo em seu texto.

As citações têm uma relação direta com as condições de produção do trabalho científico, técnico ou artístico. Escolhe-se de maneira simbólica as passagens dos autores para ajudar a refletir sobre a questão proposta, para legitimar o pensamento, para dar segurança em passagens polêmicas, por questões ideológicas, de afinidades com determinados grupos de autores que têm posições teóricas semelhantes às nossas, entre outras. Como a célebre frase de Newton afirma, temos de subir em ombros de gigantes para poder enxergar mais longe.

Ao se ler vários autores que tratam um tema específico e suas relações com outros temas, constantemente depara-se com os mesmos autores e, mais ainda, com as mesmas citações. Então, a impressão que se tem é que, no mínimo, tal autor é bastante considerado dentro do universo

pesquisado e que as passagens citadas se cristalizam como uma representação do conhecimento em sua área de estudo. As citações são, efetivamente, representações do pensamento do autor citado, que também possuem uma representação simbólica de algo para o citante.

Citações nos remetem à proposição de Karl Bühler (citado por HABERMAS, 1990, p. 78) sobre a função de representação como enunciados utilizados de modo comunicativo que expressam simultaneamente as “intenções (ou experiências) de um falante, para representar estados de coisas (ou algo que aparece no mundo) e para contrair relações com o destinatário”, ou seja, entender-se com um outro sobre algo.

#### **2.4 Análise de citação: o que podemos descobrir?**

Tamanha foi a importância do *Science Citation Index* (SCI), publicado pelo *Institute for Scientific Information* (ISI), que MacRoberts e MacRoberts (1989) afirmam que o SCI foi o marco inicial da análise de citações, em 1961.

Cronin faz uso das palavras de Morman ao fazer a definição de análise de citações:

Análise de citações na história da ciência resulta da convergência das necessidades dos ‘policy makers’ da era pós-guerra, da maturação da prática da citação bibliográfica na literatura científica durante este século, da disponibilidade do processamento eletrônico dos dados e da escola da sociologia interessada na dinâmica interna da comunidade científica. (MORMAN, 1981, p. 12).

Segundo Sengupta (1992) a análise de citações é baseada na hipótese de que todo ato de citar o autor de um trabalho prévio tem sempre algum significado. A partir da análise de citações, é possível estabelecer uma relação de comunicação que vai se constituir nas representações de certos campos do saber. A pessoa citante é responsável pelo fluxo de informação e de comunicação entre aquele que cita e o que o lê. Dessa forma, a citação proporciona a dinâmica do texto sempre vivo e circulante. É ela que impede que o texto se vá com a partida de seu autor, seja ela oral ou verbal, é a citação que mantém viva a criação de um autor.

Mesmo com todas as possibilidades que a análise de citação permite descobrir, há certos problemas que merecem ser destacados. MacRoberts e MacRoberts (1989) destacam todos as dificuldades que se encontra ao se analisar as citações dos documentos. A primeira delas é o caso

das influências formais que não são citadas, pois, segundo os autores, apenas 30% dos autores que influenciaram o trabalho de um determinado autor estão presentes na lista bibliográfica. O próximo problema é o caso das citações tendenciosas, uma vez que, segundo os autores, 38% das referências são para a fonte secundária e, assim, o crédito é tomado do verdadeiro descobridor e dado a outra pessoa. Uma terceira dificuldade diz respeito a influências informais que não são citadas, onde se enquadra todo o conhecimento tácito dos autores. MacRoberts e MacRoberts dizem que um quarto problema é sobre as auto-citações, pois elas ocorrem de forma excessiva, cerca de 10 a 30% das referências. Um outro problema é sobre os diferentes tipos de citação, uma vez que as citações afirmativas e negativas não são tratadas de forma diferenciada pelos analistas de citação, pois, segundo os autores, a simples contagem das citações implica na perda da complexidade dos eventos. O sexto problema relatado é sobre a variação na taxa de citações com diferentes tipos de publicações, nacionalidades, períodos de tempo e tamanho e tipo da especialidade, pois todos esses fatores influenciam muito nos resultados das análises. O sétimo e último problema relatado é sobre as limitações técnicas para a análise de citações, o que ocorre ainda hoje mesmo após 16 anos do texto de MacRoberts e MacRoberts. Essas limitações são referentes a: autores múltiplos, sobre a quem deve ser dado o crédito; sinônimos, para os nomes diferentes mas que dizem respeito à mesma pessoa, por exemplo I. LEAL e I. C. LEAL; homônimos, sobre a dificuldade de como distinguir indivíduos diferentes que possuem os nomes iguais; erros “clerical”, pois de 10 a 50% das referências já se encontram com algum erro na própria fonte primária; e, por fim, a cobertura da literatura, uma vez que o principal índice de citação da comunidade internacional, o SCI, por exemplo, cobre apenas 10% da literatura científica mundial.

Segundo MacRoberts e MacRoberts, nenhum desses problemas foi estudado exhaustivamente, o que abre uma gama de frentes de pesquisa na área de análise de citações.

Nesse contexto se insere o conceito de bibliometria, importante área de estudo da Ciência da Informação onde a análise de citações é de suma importância. A palavra bibliometria vem das raízes gregas *Byblos*, que significa livro e *Metricus*, que significa a ciência de medir. O termo foi cunhado por Pritchard em 1969, mas a técnica já era praticada desde 1890, conhecida como “bibliografia estatística”. Segundo Pritchard (1969), bibliometria é

a aplicação de métodos matemáticos e estatísticos para livros e outras formas de comunicação escrita, com o objetivo de trazer à luz o processo de comunicação escrita, a natureza e curso do desenvolvimento de uma maneira descritiva de contagem e análise das várias facetas da comunicação escrita. (PRITCHARD, 1969).

Sengupta apresentou em 1992 as áreas de estudo onde a bibliometria pode ser aplicada, ou seja, as áreas de aplicação da análise de citações, como pode ser visto na Tabela 2.

1. Identificação de frentes de pesquisa e crescimento do conhecimento nas diferentes disciplinas;
2. Estimação compreensiva dos periódicos secundários;
3. Identificação de usuários de diferentes assuntos;
4. Identificação de autoria e suas pesquisas em diferentes assuntos;
5. Medição da utilidade;
6. Previsão de tendências nas publicações do passado, presente e futuro;
7. Desenvolvimento de modelos experimentais;
8. Identificação dos principais periódicos nas diferentes disciplinas;
9. Formulação de uma política de aquisição baseada nas necessidades acurada;
10. Adaptação de uma política acurada;
11. Início efetivo de sistemas em rede multinível;
12. Regulação do fluxo da informação e comunicação;
13. Estudo da obsolescência e dispersão da literatura científica (agrupamento e acoplamento de artigos científicos);
14. Previsão da produtividade de editores, autores individuais, organizações, países ou de uma disciplina como um todo;
15. Projeto de uma linguagem automática para o processamento de uma auto-indexação, auto-resumo e auto-classificação; e
16. Desenvolvimento de normas de padronização.

**Tabela 2 – Áreas de aplicação da Bibliometria (SENGUPTA, 1992)**

A partir de análises de citações, formando o instrumental metodológico da bibliometria, tem-se uma das formas de se inferir o impacto da produção científica na comunidade. As citações permitem analisar quem influencia quem e determinar quais são os autores que contribuem efetivamente para o curso evolutivo da ciência. Com esse desenho, esse mapa, essa rede de conexões entre os trabalhos citantes e os citados, obtém-se uma forma de se avaliar a produção científica de um determinado assunto, região geográfica e até de um campo do conhecimento.

Cronin afirma que

é através das citações, em respeito ao crescimento do conhecimento científico, que é dada expressão substancial para o processo de inovação e, se corretamente ordenadas, podem prover o pesquisador com uma ferramenta analítica de poder sedutor e versatilidade. Os dados objetivos das citações têm sido explorados em uma variedade de aplicações, o que tem levado alguns sociólogos a descrevê-la como uma máquina de sensor de qualidade, podendo ser usado em vários graus de confiança, para estimar a qualidade, impacto, originalidade, penetração ou visibilidade da performance individual ou corporativa nas disciplinas. (CRONIN, 1984, p. 25).

Através dessa análise, podem-se extrair várias informações relevantes para a comunidade científica. Por exemplo, a tentativa de inferir quais são os pesquisadores mais importantes para um determinado campo científico, analisando-se o número de trabalhos publicados por eles e também analisando o número de citações que são feitas a eles, indicando, assim, se o impacto deles na comunidade foi bastante forte, podendo até reverter o rumo de pesquisas futuras.

A partir desses pensamentos, vislumbra-se uma forma de se representar toda a comunicação científica através das citações dos trabalhos publicados pelos autores. Com o mapa da produção científica, as agências de fomento poderiam ter uma fundamentação mais concreta ao negar ou aceitar determinadas solicitações de investimentos, pois, dessa forma, se conheceriam os pesquisadores de maior valor agregado, como dito no parágrafo anterior.

E, por fim, com esse mapa da produção científica em mãos, pode-se vislumbrar um acompanhamento maior da pesquisa nacional, o que pode ser chamado de um monitoramento da pesquisa científica nacional.

## 2.5 Índices de citação: uma ferramenta de monitoração da produção científica

Há uma grande confusão com os termos catalogação, classificação e indexação, como afirma Lancaster (1993). O autor afirma que indexação é uma expressão usada de modo impreciso, que se refere à representação do conteúdo temático de partes de itens bibliográficos completos. Wooster (1964), é bastante pragmático ao se referir à indexação como uma atribuição de termos “provavelmente relacionados de alguma forma com o conteúdo intelectual do documento original, para ajudar você a encontrá-lo quando precisar”.

Após a apreensão do conceito de indexação, pode-se agora falar do produto. Assim, índices são resultados dos processos de indexação. Conforme definido no dicionário brasileiro da língua portuguesa (FERREIRA, 1986, p. 937):

Índice. [Do lat. indice] S. m. 1. Edit. Lista detalhada dos assuntos, nomes de pessoas, nomes geográficos, acontecimentos, etc. (entradas), ordenados normalmente por ordem alfabética, com indicação de sua localização na publicação em que aparecem; índice analítico [Cf. sumário; dim. Irreg.: indículo]. 2. O conjunto de indicações impressas, especialmente de letras, feitas num livro ou num caderno para facilitar-lhe o manuseio, permitindo a localização ou o registro do assunto desejado.

Outra definição, mais específica, mas condizente com o significado indicado pelo dicionário, pode ser encontrada no texto da ISO 999:

Índice de uma publicação é uma listagem detalhada, arranjada alfabeticamente ou com um ordem classificada ou cronológica de todos os assuntos, nomes de pessoas, nomes geográficos ou de lugares, eventos e outros com indicação do seu lugar na publicação (ISO, 1996, p. 1).

Sendo esta definição especificamente para índices de uma publicação, como aqueles no final de um livro, pode-se extrapolá-lo para todos os tipos de índices, pois cada entrada em um índice serve como um ponteiro indicando onde encontrar a informação ou o documento desejados.

Nesta mesma norma da ISO (1996), são definidos alguns tipos de índices mais comumente pesquisados e construídos: índices de assuntos, índices de autores, índices de pessoas, índices de nomes de lugar, índices de abreviações, siglas, símbolos, índices de citação e índices de títulos de trabalhos.

Como esta última definição foi retirada de uma norma da ISO, existem outras várias sobre o mesmo assunto. Podemos citar, por exemplo, a norma ISO 5963, do ano de 1985, com o título *Documentation – Methods for examining documents, determining their subjects and selecting indexing terms*. Existe, também, uma norma nacional da ABNT de código NBR 6034, com o título *Preparação de índice de publicações*.

Ao trabalharmos com índices de citação, primeiramente, torna-se necessário fazer uma clara definição do termo citação. Para tanto, utilizamos a definição feita por Garfield:

Citações, geralmente compostas pelo título, autor, onde e quando publicado, provêm um precedente para ilustrar em que o autor tem a dizer. Citações são ligações formais, explícitas entre artigos que têm pontos particulares em comum. (GARFIELD, 1979, p. 1)

Uma outra definição feita também por Garfield diz que “a citação é a representação precisa, sem ambigüidade do assunto [do documento] que não requer interpretação e é imune a mudanças de terminologia” (GARFIELD, 1979, p. 3).

A partir de análises de citações, formando o instrumental metodológico da bibliometria, tem-se uma das formas de se inferir o impacto da produção científica na comunidade. As citações permitem analisar quem influencia quem e determinar quais são os autores que contribuem efetivamente para o curso evolutivo da ciência.

Ziman afirma que é interessante estudarmos o que ele chama de “árvores de citação”, observando como os trabalhos se relacionam uns com os outros, de maneira que cada um forma um nodo de uma rede, de onde partem ramificações que os ligam a vários trabalhos anteriores e posteriores. Pode-se, a partir daí, enumerar todos os escritos referentes a um trabalho específico anterior (ZIMAN, 1979).

Outro aspecto interessante de um índice de citação, conforme discutido por Garfield (1996), é a questão da avaliação da produção científica. Segundo ele, um índice de citação pode ajudar a prevenir duplicação de pesquisas, já que há cerca de 25% de pesquisas duplicadas, conforme um estudo realizado em 1963<sup>4</sup> indicou. Outro estudo realizado em 1977 por Julie Virgo, demonstrou que é bem verdadeira a correlação da análise de citações e do julgamento de líderes de frentes de pesquisa sobre câncer. O propósito do estudo foi o desenvolvimento, usando critérios objetivos,

---

<sup>4</sup> Estudo realizado em 1963 por John Martyn publicado em *New Scientist*: “*Unintentional Duplication of Research*”

de um procedimento estatístico para avaliar a importância de artigos de periódicos, sendo que o experimento mostrou que a frequência de citações foi capaz de prever consistentemente o artigo mais importante.

Um “ancestral” dos índices de citação, ou, podemos dizer, uma das primeiras aplicações desse conceito foi feita em *Shepard's Citations* (ADAIR, 1955), uma ferramenta de referência da área de Direito que tem sido utilizada desde 1873. A *Shepard's Citations* deve sua existência ao fato da Lei Americana operar sob a doutrina de *Stare Decisis*. *Stare Decisis* significa que todos os tribunais devem seguir seus precedentes, suas jurisprudências, assim como aqueles estabelecidos por tribunais de níveis mais altos. Exemplificando, um advogado deve basear seus argumentos em decisões prévias preservando um ponto de vista similar. Antes de apresentar uma decisão prévia como um precedente, o advogado deve ter a certeza de que a decisão não foi anulada, revogada ou limitada de alguma forma. Dessa forma, *Shepard's Citations* pode tornar os advogados capazes de fazerem isso com o mínimo de dificuldade.

Pode-se afirmar que o ISI tem participado da história evolutiva dos índices de citação, uma vez que o primeiro índice publicado oficialmente foi desenvolvido por sua equipe e até hoje seus índices de citação são muito utilizados e têm sido objetos de estudo de vários pesquisadores da área de citações. Nas seções seguintes apresentamos informações relevantes deste renomado centro de pesquisa e de um de seus maiores representantes, Eugene Garfield.

### **2.5.1 *Institute for Scientific Information (ISI)***

O *Institute for Scientific Information (ISI)* foi fundado em 1958 por Eugene Garfield, hoje com o nome de Thomson ISI, segundo informações do *site* do instituto, disponível no endereço <http://www.isinet.com/isi/about>. Segundo o *site*, o ISI tem provido mais de sete milhões de pesquisadores com seus produtos e serviços que os ajudam a transformar o processo de descoberta, possibilitando essa comunidade a obter acesso a pesquisas históricas e às mais recentes descobertas.

O instituto tem mais de 850 empregados ao redor do mundo, com escritórios nos Estados Unidos, Reino Unido, Irlanda, Tóquio e Singapura e, hoje em dia, o ISI faz parte do *The Thomson Corporation*, cuja aquisição foi realizada em 1992. Essa corporação é líder mundial em soluções



de informações integradas para negócios, com renda de US\$ 7,8 bilhões em 2002, provendo informações de valor agregado, *softwares* e ferramentas para mais de 20 milhões de usuários.

O ISI publica anualmente os três índices de citação mais conhecidos da literatura: *Science Citation Index* (SCI), *Social Science Citation Index* (SSCI) e *Arts and Humanities Citation Index* (AHCI). Segue abaixo uma descrição sucinta de cada índice:

*Science Citation Index* (SCI) <sup>5</sup>:

É um índice multidisciplinar, cobrindo mais de 5900 periódicos sobre 164 disciplinas. Este índice provê o acesso a informações desde o ano de 1945 até hoje, havendo uma média de 17750 novos registros por semana e contem mais de 17 milhões de registros. Para exemplificar algumas disciplinas que são cobertas por este índice, podemos citar: Agricultura, Astronomia, Bioquímica, Biologia, Biotecnologia, Química, Ciência da Computação, Ciências Materiais, Matemática, Medicina, Neurociência, Oncologia, Pediatria, Farmacologia, Física, Psiquiatria, Cirurgia, Veterinária e Zoologia.

*Social Science Citation Index* (SSCI) <sup>6</sup>:

É também um índice multidisciplinar que abrange os periódicos das ciências sociais. O índice é composto por cerca de 1725 periódicos, cobrindo mais de 50 disciplinas, provendo acesso a publicações desde 1956. São adicionados cerca de 2700 registros por semana e são indexados mais de 3,15 milhões de registros. Algumas disciplinas que cobertas pelo SSCI são: Antropologia, História, Relações Industriais, Ciência da Informação, Biblioteconomia, Direito, Linguística, Filosofia, Psicologia, Psiquiatria, Ciências Políticas, Saúde Pública, Serviço Social, Sociologia, Estudos Urbanos e Estudos da Mulher.

*Arts and Humanities Citation Index* (AHCI) <sup>7</sup>:

Este índice cobre as publicações de mais de 1144 periódicos líderes mundiais sobre artes. São indexados artigos publicados nestes periódicos desde 1975, sendo adicionados, em média, 2300 novos registros por semana, sendo que contém, atualmente, cerca de 2,5 milhões de registros. Abrange publicações das áreas de Arqueologia, Arquitetura, Arte,

---

<sup>5</sup> Informações disponíveis em <http://www.isinet.com/isi/products/citation/scie/>, acessado em 13/08/2003.

<sup>6</sup> Informações disponíveis em <http://www.isinet.com/isi/products/citation/ssci/>, acessado em 13/08/2003.

<sup>7</sup> Informações disponíveis em <http://www.isinet.com/isi/products/citation/ahci/>, acessado em 13/08/2003.

Estudos Asiáticos, Dança, Folclore, História, Linguagem, Lingüística, Revisões Literárias, Literatura, Música, Filosofia, Poesia, Rádio, Televisão e Cinema, Religião e Teatro.

Um fator que confirma a necessidade de uma ferramenta como a apresentada neste trabalho, é o fato que institutos internacionais de produção de bases de dados da produção científica, como o ISI, não se constituem em fontes exaustivas e muito menos representam a produção científica nacional, bem como dos demais países do terceiro mundo, conforme afirma Alvarenga (1996).

Analisando o caso brasileiro, Castro (apud FREITAS, 1990) afirma que os bancos de dados do ISI não refletem nossa produção registrada em literatura periódica, uma vez que, em 1982, apenas 14% da produção brasileira foi publicada em canais internacionais e que 3000 periódicos rastreados pelo ISI, apenas quatro teriam sido publicados no Brasil.

### 2.5.2 *Eugene Garfield*

Eugene Garfield é o nome proeminente da área de pesquisa de citação, despontando como um dos maiores pesquisadores e conhecedores do assunto. Garfield foi o fundador do ISI e, já tendo sido presidente do ISI, após sua venda, hoje é o presidente de honra do mesmo instituto.

O interesse de Garfield, como descrito por ele mesmo em (GARFIELD, 1998), pela área de pesquisa de índices de citação se inicia a partir de sua investigação na área de lingüística, nas características de indexação de artigos científicos e ao deparar-se com o Shepard's Citations. Seu conhecimento do *Shepard's Citation* se dá através do vice-presidente aposentado da instituição que trabalhava nele, W. C. Adair, no ano de 1953, quando este convidou Garfield para fazer parte do *Johns Hopkins Welch Medical Indexing Project*. Ao encerramento deste projeto, em junho de 1953, Garfield ingressou-se na *Columbia University School of Library Science*, daí, em 1954, ele escreve seu primeiro artigo propondo a criação de um índice de citação, publicado na conceituada revista *Science* em 1955 (GARFIELD, 1955). Seu propósito principal era o de melhorar a recuperação de informação científica.

Em 1958, no *International Conference on Scientific Information*, Garfield propõe um índice multidisciplinar, unificado para a ciência, que indexa todos os jornais e periódicos, citando, pela primeira vez, o que seria o *Science Citation Index* (SCI).

Conforme descreve Garfield (1998), o SCI foi proposto numa época ideal, pois:

Naquele atual momento, havia uma insatisfação generalizada com as tradicionais indexações orientadas a disciplinas e com os serviços de resumos, pois a indexação era excessivamente demorada, inconsistente e desordenada. As políticas de seleção deixaram grandes gaps em sua cobertura.

Após todo seu sucesso na área de índices de citação, colaborando na criação dos três índices publicados pelo ISI, (SCI, SSCI e AHCI), hoje Garfield é presidente e editor da revista *The Scientist*, presidente de honra do ISI, membro do Comitê do *John Scott Medal*. Vale ressaltar que ele foi presidente do ASIS&T (*American Society for Information Science & Technology*), membro do *Overseers Libraries of University of Pennsylvania* e diretor e membro do grupo de pesquisa *Research America*.

## **2.6 Utilização de indicadores científicos no apoio à tomada de decisão**

Após as grandes guerras, a partir da década de 1960, os interesses se voltaram para a investigação científica e tecnológica com maior intensidade, dando ênfase também à coleta de informações e estatísticas sobre essas atividades. Essa corrente se manifestou principalmente nos países economicamente mais desenvolvidos e, no âmbito dos órgãos internacionais, a Unesco teve liderança, cunhando o conceito de “potencial científico e tecnológico nacional”.

No início dessas atividades, até meados da década de 1970, as iniciativas se baseavam em coletar e produzir informações e estatísticas relacionadas aos insumos ou *inputs* alocados para as atividades de C&T. Após esse período, as informações sobre os produtos ou *outputs* das atividades de C&T começavam a ser compiladas como subprodutos da Tecnologia da Informação e ainda não eram utilizados para as políticas de C&T. Isso até perceberem que era de grande valia a utilização de informações quantitativas para colaborar com o planejamento, monitoramento e avaliação da ciência, quando a ciência e tecnologia foram removidas da periferia das políticas governamentais ocupando uma posição central (VELHO, 2001).

Com suporte de números que ajudassem nas políticas governamentais para o desenvolvimento da ciência de uma nação, a burocracia estatal responsável por essas políticas foi fortalecida em todos os países. Conseqüentemente, Velho afirma que “a administração pública começou a ter maior

interesse nos resultados de P&D e na sua disseminação, dando lugar ao aparecimento de serviços técnicos e científicos, entre os quais o de compilação de estatísticas de C&T para a construção de indicadores.” (VELHO, 2001, p. 111). Nesse momento, os custos das atividades de pesquisa científica estavam se tornando cada vez maiores e competindo com os demais setores de investimento público. A autora conclui que

[...] a ênfase nos indicadores quantitativos parece derivar da premissa de que eles proporcionariam uma base racional completamente objetiva para o planejamento em C&T e de que uma ciência da ciência mecanicista não era apenas necessária, mas também suficiente (VELHO, 2001, p. 111).

Velho (2001) utiliza outro argumento para embasar a necessidade de se ter em mãos dados quantitativos da produção técnica e científica de um país. Segundo ela, devido à globalização da economia e à crescente competitividade entre nações e empresas, há benefícios potenciais para o país que consiga entender sua posição em relação aos seus competidores em diferentes áreas da ciência, para que, com isso, possam explorar as melhores oportunidades que possam surgir.

A aplicação de métodos de medida da produção científica para sua avaliação é efetuada sob algumas premissas, segundo Rousseau (1998). O autor afirma que: 1) o progresso científico é alcançado mediante o trabalho de cientistas; 2) esses cientistas constroem seus trabalhos sobre as obras de colegas e precursores de sua área; 3) os resultados desse trabalho são publicados; sendo, portanto, 4) submetidos à apreciação dos pares. 5) A próxima afirma que em suas publicações, os cientistas mostram como construíram seus trabalhos a partir de obras anteriores mencionando-as em seus textos em uma lista de referências. Como relatado neste trabalho sobre a importância e a necessidade da comunicação científica para o progresso da mesma, 6) as revistas científicas desempenham papel essencial nesse processo de comunicação. 7) A seguinte diz que o número de publicações de um grupo de pesquisa pode ser considerado como um indicador de sua produção científica, uma outra forma de mensuração é 8) o número de vezes que as obras de um grupo são citadas por outras publicações, útil para a medida do impacto ou da visibilidade internacional dessas obras. 9) Rousseau afirma que é possível distinguir as revistas “centrais”, mais importantes, das periféricas e, por fim, diz que 10) o *Institute for Scientific Information* (ISI) cobre a maioria das revistas internacionais importantes, através de seus índices de citação.

### **3 Metodologia**

O presente trabalho pode ser entendido como uma pesquisa aplicada, pois visa uma utilização imediata dos conhecimentos produzidos (CHIZZOTTI, 1998, p. 35) e caracteriza-se pelo interesse prático, onde seus resultados são aplicados e utilizados imediatamente (MARCONI; LAKATOS, 1988, p. 19). No presente caso, isso se traduz na construção de uma ferramenta, um produto ou ainda um sistema de computador.

A metodologia utilizada no desenvolvimento deste trabalho é explicitada em dois momentos, assim como os objetivos apresentados logo atrás, a fim de esclarecer como cada um deles foi atingido.

#### **3.1 Etapa de implementação**

Para alcançar o objetivo da primeira parte do trabalho, que consiste no desenvolvimento da aplicação para a análise das citações presentes nas teses e dissertações defendidas e aprovadas no PPGCI, foram privilegiadas ferramentas gratuitas, preferencialmente softwares livres e de código aberto.

O trabalho desenvolvido se baseou na plataforma Dspace de criação de bibliotecas digitais, a fim de agregar um novo módulo a uma ferramenta já conhecida e utilizada pelas comunidades científicas. A idéia inicial era de utilizar o software Greenstone<sup>8</sup>, para a construção da biblioteca digital, onde seria o suporte para a aplicação da análise das citações. Com o conhecimento do Dspace, por motivos tecnológicos e por ser uma ferramenta utilizada por grandes universidades internacionais, como *Columbia University*, *Cornell University*, *Massachusetts Institute of Technology*, *Ohio State University*, *University of Cambridge*, *University of Rochester*, *University of Toronto*, *University of Washington*, sendo uma ferramenta bastante promissora, foi alterado o núcleo do trabalho, pois todo o desenvolvimento deveria seguir os padrões da plataforma selecionada.

---

<sup>8</sup> Software de gerenciamento, criação e manutenção de bibliotecas digitais, de código aberto, desenvolvido pela *New Zealand Digital Library Project*. On-line: <http://www.greenstone.org>. Acessado em 03/06/2005.

Após o contato com a Professora Ana Alice Baptista, da Universidade do Minho (UMinho), de Portugal que já utiliza o Dspace desde 20/11/2003, disponível no endereço <https://repositorium.sdum.uminho.pt/index.jsp>, teve-se acesso a uma versão do Dspace traduzido para o Português de Portugal, o que facilitou o uso, já que as versões disponíveis pelo site do Dspace são em Inglês. Para iniciar o trabalho, foram feitas algumas alterações na versão da UMinho, para a customização do Dspace para a ECI-UFMG, assim como mudanças dos links principais para os *sites* da UFMG e substituição dos logotipos da UMinho pelos da ECI e da UFMG.

Como o Dspace é um software livre e sua licença permite que se altere a ferramenta, o desenvolvimento foi feito utilizando-se as mesmas ferramentas indicadas pela HP e MIT: linguagem de programação Java<sup>9</sup>, banco de dados PostgreSQL<sup>10</sup>, servidor Web Tomcat<sup>11</sup>, distribuído pela Apache Foundation. Foi ainda utilizada a ferramenta de desenvolvimento Eclipse<sup>12</sup> e tudo sendo executado sobre o sistema operacional Linux Slackware<sup>13</sup>.

Para efetivar a análise de citações proposta, foram selecionados alguns indicadores, ou variáveis, componentes de uma citação. Os valores dados aos conceitos selecionados, transformando-os em variáveis, são quantidades que se alteram em cada caso particular (MARCONI; LAKATOS, 1988, p. 151). Os indicadores trabalhados podem ser divididos em dois grupos, como relacionados nas próximas subseções.

### **3.1.1 Variáveis independentes**

As variáveis independentes são aquelas “que influenciam, determinam ou afetam uma outra variável, sendo fator determinante, condição ou causa para determinado resultado, efeito ou consequência” (MARCONI e LAKATOS, 1988, p. 163). Foram selecionadas para este trabalho

---

<sup>9</sup> Uma das linguagens de programação mais utilizadas na atualidade. Informações disponíveis em <http://java.sun.com>. Acessado em 03/06/2005.

<sup>10</sup> Banco de dados relacional, de código aberto, disponível apenas para sistema operacional Linux. Maiores informações disponíveis em <http://www.postgresql.org>. Acessado em 03/06/2005.

<sup>11</sup> Servidor Tomcat, um software livre e gratuito. Informações em <http://jakarta.apache.org/tomcat>. Acessado em 03/06/2005.

<sup>12</sup> Ferramenta para desenvolvimento de sistemas na linguagem de programação Java, de código aberto. Maiores informações em <http://www.eclipse.org>. Acessado em 03/06/2005.

<sup>13</sup> Versão de distribuição do sistema operacional Linux, gratuita, de código aberto. Maiores informações em <http://www.slackware.com>. Acessado em 03/06/2005.

quatro variáveis: Linha de Pesquisa do programa de Pós-Graduação, orientador, único documento, ao se analisar as referências bibliográficas de apenas uma tese ou dissertação e geral, contemplando todos os documentos do universo trabalhado. A partir da variável de ambiente se inicia a análise das referências, primeiro passo para a seleção do grupo de documentos analisados.

Abaixo segue uma tabela com a lista das variáveis independentes selecionadas.

<b>Variáveis independentes</b>
Linha de Pesquisa
Orientador
Documento único
Geral

**Tabela 3 – Variáveis independentes**

As variáveis independentes são cadastradas quando da adição do documento à biblioteca digital do Dspace. As inclusões foram feitas com a interface própria do Dspace, utilizando os metadados no padrão Dublin Core, segundo o mapeamento utilizado pelo Dspace:

<b>Informação da tese ou dissertação</b>	<b>Metadado no padrão Dublin Core</b>
Autor	contributor.author
Data da defesa	date.issued
Idioma	language.iso
Linha de Pesquisa	subject.classification
Orientador	contributor.advisor
Palavras-chave	subject
Referência bibliográfica	relation
Resumo	description.abstract
Resumo em Inglês ( <i>Abstract</i> )	description.abstract
Tipo de documento (dissertação de mestrado ou tese de doutorado)	type
Título	title

**Tabela 4 – Mapeamento: informações do documento x metadado Dublin Core**

### 3.1.2 Variáveis dependentes

O segundo grupo de variáveis, denominadas variáveis dependentes, é composto por “aqueles valores, fenômenos ou fatores a serem explicados ou descobertos, em virtude de serem influenciados, determinados ou afetados pela variável independente” (MARCONI e LAKATOS, 1988, p. 163).

As variáveis dependentes foram extraídas das citações, formando um grupo de 7 variáveis: autor, formado por todos os autores do conjunto de citações; ano de publicação, formado pelos anos das citações analisadas; tipologia da publicação, podendo assumir o valor de artigo de periódico, livro, trabalho apresentado em congresso, dissertação, tese, norma técnica, nota de aula, legislação ou uma última, não identificada, quando a citação não se enquadrar em nenhuma das categorias anteriores; país de publicação, podendo ser qualquer país do globo terrestre; idioma, sobre a língua utilizada na citação; assunto dos títulos, formado por termos extraídos dos títulos dos documentos analisados, não sendo utilizado quando a variável de ambiente for de um documento apenas; e assunto das referências, termo extraído dos títulos das referências dos documentos analisados.

Abaixo segue uma tabela com a lista das variáveis dependentes selecionadas.

<b>Variáveis dependentes</b>
Autor
Ano
Tipologia
País
Idioma
Assunto dos títulos
Assunto das referências

**Tabela 5 – Variáveis dependentes**



### 3.1.3 Fases da implementação

Para a contagem dos indicadores foi acrescentada ao Dspace uma nova funcionalidade, a de cadastro das informações extraídas de cada referência de cada documento. Para tanto, a partir da escolha do usuário pelo documento a ser trabalhado, todas as referências do documento selecionado eram exibidas com campos de formulário para o cadastro das sete variáveis dependentes da referência, como o formulário apresentado no Anexo 7. Foram criadas estruturas no banco de dados do Dspace próprias para o armazenamento desses indicadores.

As informações de ano, tipo de publicação, idioma e país não necessitaram de ajustes após o seu cadastramento. Já, os nomes dos autores, uma das mais interessantes a trabalhar, precisavam de uma segunda manipulação, uma vez que era evidente que o número de autores não significava que eram autores diferentes. Vários autores apareciam mais de uma vez e vários nomes escritos de formas diferentes eram encontrados para designar o mesmo autor. Por esse motivo, foi trabalhado o conceito de arquivo de autoridade, do Inglês *authority file*.

Segundo Fast, Leise e Steckel (2005), arquivo de autoridade é uma ferramenta simples que possibilita preencher a lacuna entre a linguagem natural e complexos vocabulários controlados. Com o arquivo de autoridade, ao invés de todos os termos sinônimos serem tratados como termos iguais, um é identificado como o termo preferido e os outros são considerados termos variantes, definindo qual termo é o cabeçalho e qual é sua referência cruzada.

Por exemplo, o autor *LEAL, Igor Campos* poderia ser encontrado em outra referência como *LEAL, Igor*, ou *LEAL, I C*, ou ainda *LEAL, Igor C*. Seguindo o padrão utilizado por outros sistemas de análise de citações, como o ISI, com o *Web of Science* e seus índices de citação, o Citeseer e o Portal ACM, a forma canônica selecionada, ou seja, aquela que poderia representar todos os outros possíveis nomes, foi o último sobrenome em letras maiúsculas seguido das iniciais de cada nome. No exemplo dado acima, a forma canônica deste autor seria a opção *LEAL, I C*.

Com isso, surgiu a necessidade de descobrir dentre todos os autores das referências aqueles em que seriam os possíveis representantes da sua forma canônica. Foi utilizado um algoritmo de casamento de padrão aproximado, muito conhecido na área da Ciência da Computação, denominado Shift-And-Aproximado (ZIVIANI, 2004, p. 321), algoritmo que descobre a

existência ou não de um texto, chamado padrão, em outro. A seleção deste algoritmo seguiu a indicação de Lawrence, Giles e Bollacker, pois esse algoritmo se enquadra no que os autores chamaram de Medidas de Distância de Edição, do Inglês *Edit Distance Measures* (LAWRENCE; GILES; BOLLACKER, 1999, p. 2). O algoritmo utilizado, implementado na linguagem de programação Java, está descrito no Anexo 2.

Em poucas palavras, o algoritmo Shift-And-Aproximado tenta casar o padrão com o texto, executando ações de inclusão, substituição ou remoção de um caractere quantas vezes for necessário para transformar um texto em outro. O número máximo de vezes em que uma dessas ações é executada é um parâmetro pré-definido e, após realizado alguns testes, foi definido que para autores com nomes de 1 a 3 caracteres não seria aceita nenhuma ação, de 4 a 6 caracteres seria aceita apenas uma ação e para nomes maiores do que 6 caracteres seriam aceitas até duas ações. Essa foi a distribuição selecionada porque quanto menor o nome do autor, o número de outros nomes que casariam com ele era muito grande e apresentava um desempenho muito lento, se o número máximo de ações fosse muito grande.

Com o receio de se criar um método automático de seleção de formas canônicas e este associar nomes iguais a pessoas diferentes, foi implementado um formulário onde o usuário selecionaria a forma canônica para cada um dos autores das referências, apresentado no Anexo 8. Foram criadas estruturas no banco de dados do Dspace para armazenar as formas canônicas selecionadas para cada autor.

Para a geração dos termos dos assuntos contidos nos títulos dos documentos e nas referências foi utilizada a ferramenta ptx existente no Linux Slackware, que implementa o KWIC, sigla do termo em Inglês *Key-Word In Context*. O KWIC é uma técnica de criação de índices de assuntos rotados em que cada palavra contida no título do documento, utilizando-a como uma entrada no índice, excluindo-se palavras sem significados que interessasse, as chamadas *stop words*. Apesar de ser uma técnica bastante simples, é também eficiente, pois cada palavra-chave é vista em seu contexto (LUHN, 1959). Segue no Anexo 1 a lista de stop words utilizada neste trabalho na geração dos termos pelo KWIC.

Como aconteceu com os nomes dos autores, vários assuntos extraídos tinham grafias diferentes (masculino ou feminino, singular ou plural, línguas diferentes), mas remetiam ao mesmo conceito. Para corrigir esse problema, também foi trabalhado o conceito de arquivos de

autoridade, onde era selecionada a forma canônica de representação do assunto. A forma canônica trabalhada foi a palavra em Português, no gênero masculino e no singular.

Ao realizar testes com o algoritmo Shift-And-Aproximado com os assuntos, os arquivos de autoridades indicados eram muito grandes e continham muitos erros, uma vez que o algoritmo não faz nenhuma espécie de análise semântica dos termos, associando formação com informação, por exemplo. Para os nomes dos autores, isso não era problema pois os nomes poderiam estar com grafias diferentes e, mesmo assim, representar a mesma pessoa, como, por exemplo, *LEAL, Ygor* e *LEAL, Igor*, onde poderia ser um erro de digitação. Por isso, o método utilizado na geração dos possíveis arquivos de autoridade para os assuntos não foi o Shift-And-Aproximado. Um algoritmo bem simples de casamento de padrão foi implementado em que um assunto era similar ao outro caso o número de caracteres iniciais iguais era maior que o tamanho do menor assunto dividido por 2, para assinalar apenas as palavras com mais da metade iguais.

Como para os nomes dos autores, foi implementado um formulário de cadastro da forma canônica de cada termo gerado pelo método KWIC, como visto no Anexo 9, devido ao mesmo receio do algoritmo associar algum termo a uma forma canônica de forma incorreta, principalmente no que diz respeito a palavras em idiomas distintos. Também foram criadas estruturas no banco de dados do Dspace para armazenar as formas canônicas selecionadas para cada assunto.

Por fim, temos a parte principal do trabalho, onde as análises dos indicadores dos documentos poderiam ser vislumbradas. Foi implementada a visualização das variáveis dependentes em forma de gráficos, para que o usuário tenha maior facilidade para o acompanhamento das variações de cada indicador em cada item que o compõe. Os gráficos foram implementados utilizando a biblioteca de geração dinâmica de gráficos chamada de JFreeChart<sup>14</sup>. Para a visualização dos indicadores através desses gráficos, o usuário deve, primeiro, selecionar a variável independente desejada e em seguida são exibidos os gráficos das variáveis dependentes. Esses gráficos são:

1. Gráfico de País: gráfico de colunas, onde em um eixo do gráfico temos todos os países presentes nas referências do(s) documento(s), em ordem alfabética e, no outro, o número de referências deste país;

---

<sup>14</sup> Biblioteca na linguagem de programação Java, na versão 1.0.0 rc1, disponível em <http://www.jfree.org/jfreechart>. Acessado em 03/06/2005.

2. Gráfico de Ano: gráfico de linha, onde em um eixo do gráfico temos, a partir do ano mais antigo presente nas referências do(s) documento(s) até o último e, no outro, o número de referências deste ano;
3. Gráfico de Idiomas: gráfico de colunas, onde em um eixo do gráfico temos todos os idiomas presentes no(s) documento(s), em ordem alfabética e, no outro, o número de referências deste idioma;
4. Gráfico de Tipo de Publicação: gráfico de pizza, onde temos todos os tipos de publicação possíveis, em ordem alfabética, divididas em faixas de áreas proporcionais à sua representatividade no conjunto analisado;
5. Gráfico de Autores: gráfico de colunas, onde em um eixo do gráfico temos os 10 autores mais citados no(s) documento(s), em ordem alfabética e, no outro, o número de referências deste autor;
6. Gráfico de Assuntos dos Títulos: gráfico de colunas, onde em um eixo do gráfico temos os 10 assuntos mais utilizados nos títulos dos documentos, em ordem alfabética e, no outro, o número de vezes em que este assunto foi utilizado;
7. Gráfico de Assuntos dos Títulos das Referências: gráfico de colunas, onde em um eixo do gráfico temos os 10 assuntos mais utilizados nos títulos das referências dos documentos, em ordem alfabética e, no outro, o número de vezes em que este assunto foi utilizado.

A maior premissa para essa etapa do trabalho era a de que toda implementação deveria ser feita de forma que o impacto causado nas Instituições que já utilizassem o Dspace fosse o menor possível, caso quisessem acrescentar o módulo de análise de citações aqui apresentado. Por isso, tudo o que foi acrescentado ao Dspace foi feito seguindo exatamente o mesmo padrão de desenvolvimento, para que outras instituições possam ter acesso ao código e contribuir para sua melhoria, seguindo a mesma idéia do Dspace.

### **3.2 Etapa experimental**

O segundo momento desta dissertação se deu a partir da utilização da ferramenta desenvolvida, a fim de se realizar a parte experimental da pesquisa, onde se aplica a experimentação para

verificar a validade do fenômeno analisado, tentar explicá-lo e compreendê-lo (CHIZZOTTI, 1998). Neste caso, o experimento foi realizado com a utilização da ferramenta implementada com documentos reais. Para tanto, foram selecionadas as teses e dissertações defendidas e aprovadas pelo PPGCI da UFMG.

O primeiro passo foi a obtenção das teses e dissertações em formato PDF, com o contato direto com a secretaria do PPGCI, o que resultou em 55 documentos, sendo 47 dissertações de Mestrado e 8 teses de Doutorado, com datas de defesa variando de 26/11/1998 a 13/07/2004, não significando que todas as teses e dissertações nesse período de tempo tenham sido contempladas, sendo que dos 55 documentos apenas um foi defendido no ano de 1998, 16 em 2000, 14 em 2001, 9 em 2002, 9 em 2003 e 6 documentos em 2004. Esses documentos foram incluídos na biblioteca digital do Dspace seguindo 3 etapas, como apresentado pelos formulários dos Anexos 4, 5 e 6.

As análises também foram feitas separando as referências por Linha de Pesquisa. O PPGCI se divide em três linhas: Gestão da Informação e do Conhecimento (GIC), Informação, Cultura e Sociedade (ICS) e Organização e Tratamento da Informação (OTI). Os 55 documentos se distribuíram em 30 para a Linha de Pesquisa GIC, 18 para ICS e 7 para OTI.

O segundo momento consistiu na inclusão desses documentos na biblioteca digital do Dspace, o que foi efetuado com o auxílio de alguns alunos da Graduação da turma do segundo semestre de 2004 da disciplina Bibliotecas Digitais, ministrada pelo Professor Marcello P. Bax, em um trabalho prático. Cada referência bibliográfica encontrada nos documentos foi inserida, uma a uma, resultando em um trabalho bastante demorado, mas, por outro lado, de muito valor, pois resultou em um número total de 4.495 referências a serem analisadas.

A terceira etapa do trabalho foi o cadastro dos indicadores para cada uma das referências levantadas na etapa anterior. Foram gastos cerca de dois meses para finalizar essa fase. Depois de finalizada, ao se fazer uma verificação nas informações salvas, foram detectados alguns erros no cadastramento dos indicadores, provavelmente por ter sido um trabalho penoso e cansativo. O principal erro encontrado referiu-se aos países de publicação das referências no quesito artigos de periódicos, além de perceber determinados autores escritos de forma errada, de maneira em que não era possível a identificação de mais de um autor quando era de autoria múltipla. Mas, ainda a tempo, foram gastas mais duas semanas para sanar esses problemas, para pesquisar na Internet os países dos editoriais dos periódicos das referências identificadas e corrigir os nomes dos autores.

As referências com autoria múltipla foram consideradas para a contagem de todos os autores presentes na referência um a um. Sendo assim, o número de 4.495 prováveis autores, caso todas as referências fossem de autoria única, subiu para 5.646 autores.

O quarto momento desse trabalho diz respeito à criação dos arquivos de autoridade para os autores. Para cada um dos 5.646 autores, foi cadastrada a chamada forma canônica de representação do nome do autor. Após a seleção da forma canônica para os 5.646 autores, esse número caiu para 3.190 autores distintos.

A quinta etapa de desenvolvimento diz respeito aos procedimentos necessários para a contagem dos assuntos contidos nos títulos dos documentos e nas referências. Aplicando-se o KWIC nos 55 títulos, foram extraídos 524 assuntos, uma parte deles é apresentada no Anexo 3. Já, com a rotação nos títulos das 4.495 referências, foram encontrados 22.461 assuntos. Após a seleção da forma canônica para cada um dos 22.985 assuntos, foram obtidos 280 assuntos distintos para os títulos e 2.813 para as referências.

Para os outros indicadores que foram trabalhados, como país de publicação da referência, ano de publicação, idioma utilizado e tipo de publicação, não houve a necessidade de se fazer algum tipo de processamento sobre eles, pois a contagem foi efetuada diretamente com as informações cadastradas.

Resumidamente, segue abaixo uma lista com os números do universo trabalhado.

Número de documentos	55
Distribuição pelo tipo de documento: Dissertações de Mestrado	47
Teses de Doutorado	8
Datas das defesas	De 26/11/1998 a 13/07/2004
Distribuição pelos anos de defesa: 1998	1
2000	16
2001	14
2002	9
2003	9
2004	6
Distribuição pelas Linhas de Pesquisa: Gestão da Informação e do Conhecimento	30
Informação, Cultura e Sociedade	18
Organização e Tratamento da Informação	7
Número de referências bibliográficas	4.495
Número de autores sem arquivos de autoridade	5.646
Número de autores com arquivos de autoridade	3.190
Número de assuntos dos títulos dos documentos sem arquivos de autoridade	524
Número de assuntos dos títulos dos documentos com arquivos de autoridade	280
Número de assuntos dos títulos das referências sem arquivos de autoridade	22.461
Número de assuntos dos títulos das referências com arquivos de autoridade	2.813

**Tabela 6 – Universo de trabalho**

## ***4 Apresentação dos resultados***

Com as quatro variáveis independentes (linha de pesquisa, orientador, documento único e geral) e as sete dependentes (autor, país, idioma, ano, assunto dos títulos, assunto das referências, tipo de publicação) selecionadas, podem ser estabelecidas, no mínimo, 28 análises. Esse número poderia ser maior, caso o número de possibilidades de cada variável independente também fosse considerada nessa multiplicação, como, por exemplo, as três linhas de pesquisa do PPGCI.

A apresentação dos resultados encontrados evidencia a gama de possibilidades de análise que o usuário do sistema tem em mãos. Essas análises ainda poderiam ser mais aprofundadas, ao acrescentar mais de uma variável dependente, como, por exemplo, o número de vezes em que os assuntos aparecem no decorrer dos anos para uma linha de pesquisa qualquer. Análises desse tipo são relatadas como Trabalhos Futuros e não são abordadas neste trabalho.

As seções seguintes apresentam os resultados encontrados na parte experimental do trabalho, separados por variável independente e, a partir dessa variável, cada uma das variáveis dependentes são relatadas. Vale ressaltar que a análise para um único documento não será apresentada, uma vez que ela pode ser entendida como um detalhamento das outras três, como visto nos Anexos 10 e 11. Não cabe análise nesses resultados se apresentado apenas um documento, uma vez que, se quiséssemos alcançar tal objetivo, seria necessário apresentar os resultados dos 55 documentos um por um, pois estaríamos de posse de um conjunto completo de elementos para analisar.

### **4.1 Resultados Gerais**

De modo a obtermos um primeiro “retrato” da produção científica do PPGCI, é necessário que tenhamos em mãos informações referentes a todo universo de trabalho. Essa análise, chamada de Análise Geral, representa os resultados obtidos com todos os 55 documentos presentes na biblioteca digital.



Na análise geral, não foi realizada a análise da distribuição das citações por Assunto das Referências; essa informação não agrega valor, já que há muita diferença para cada uma das Linhas de Pesquisa.

O número médio de referências por documento foi de 81 e os resultados encontrados foram os descritos nos gráficos abaixo.

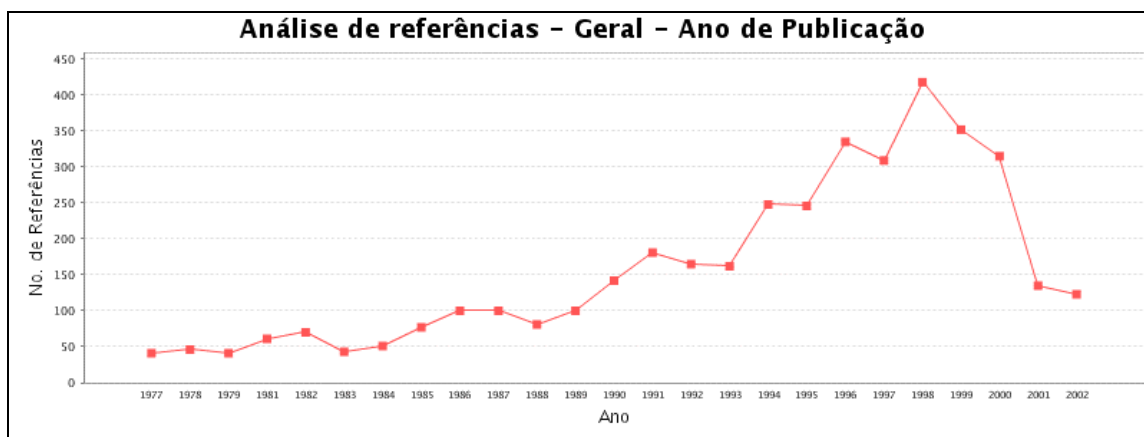


Gráfico 2 – Análise Geral das citações por Ano

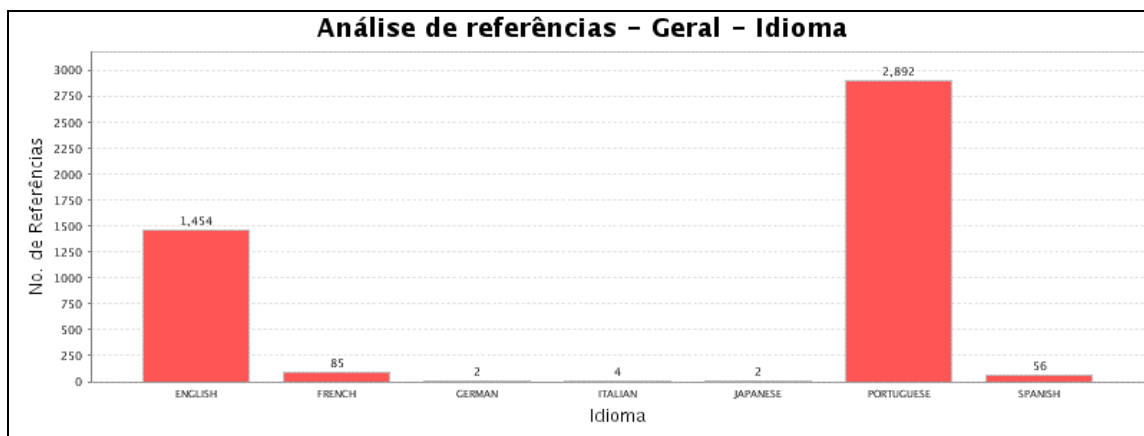


Gráfico 3 – Análise Geral das citações por Idioma

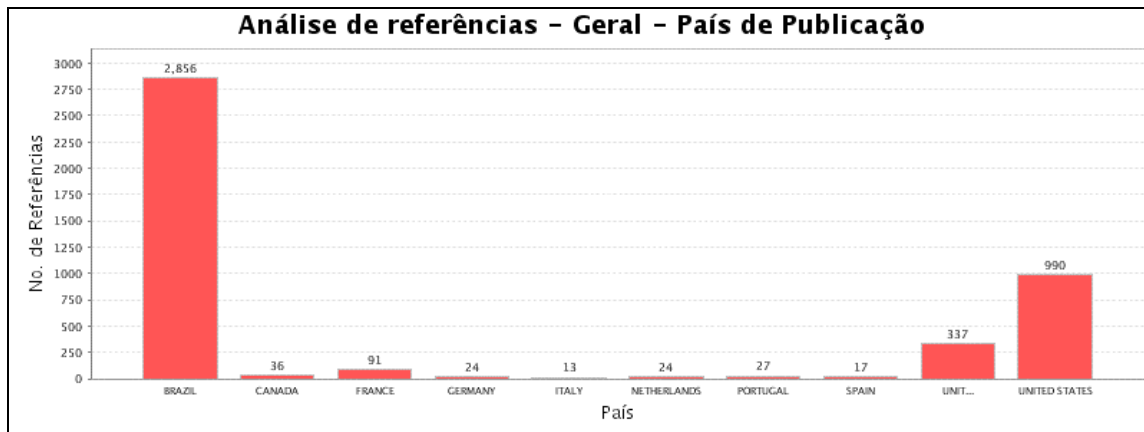


Gráfico 4 – Análise Geral das citações por País de Publicação

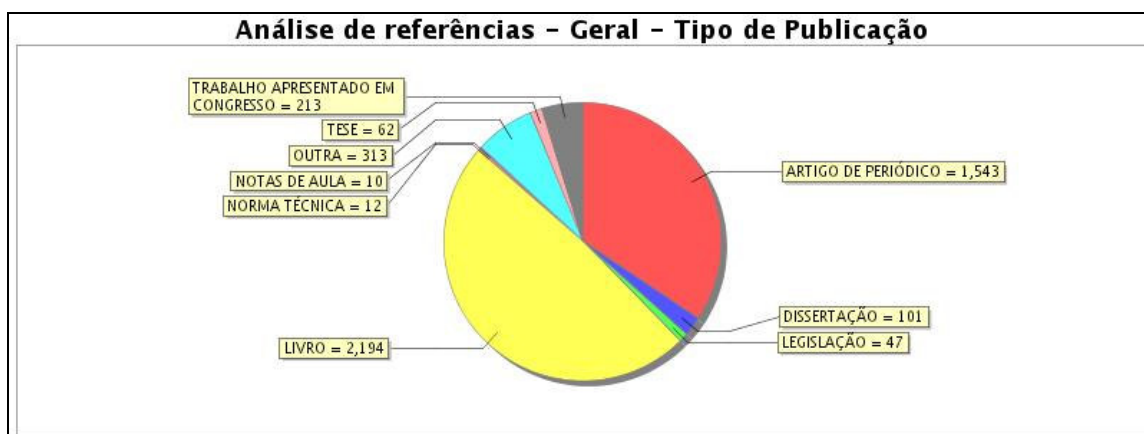


Gráfico 5 – Análise Geral das citações por Tipo de Publicação

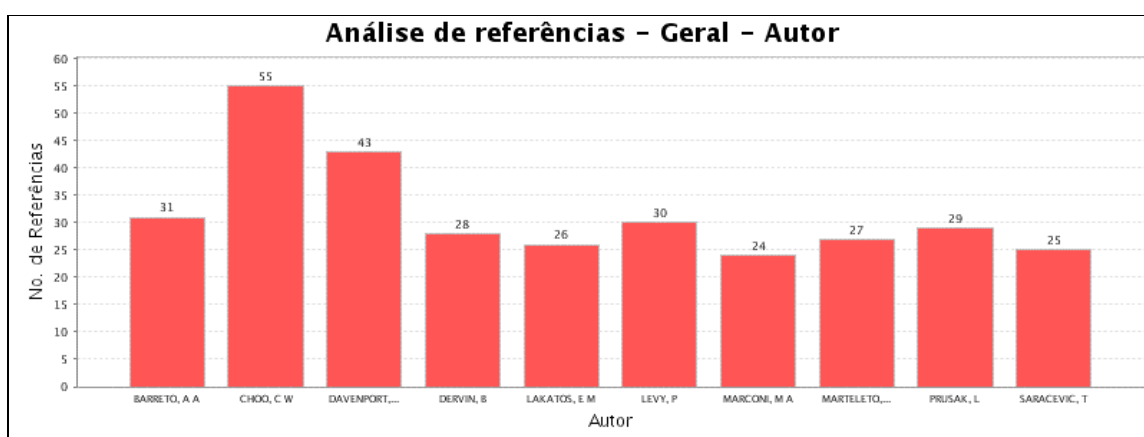
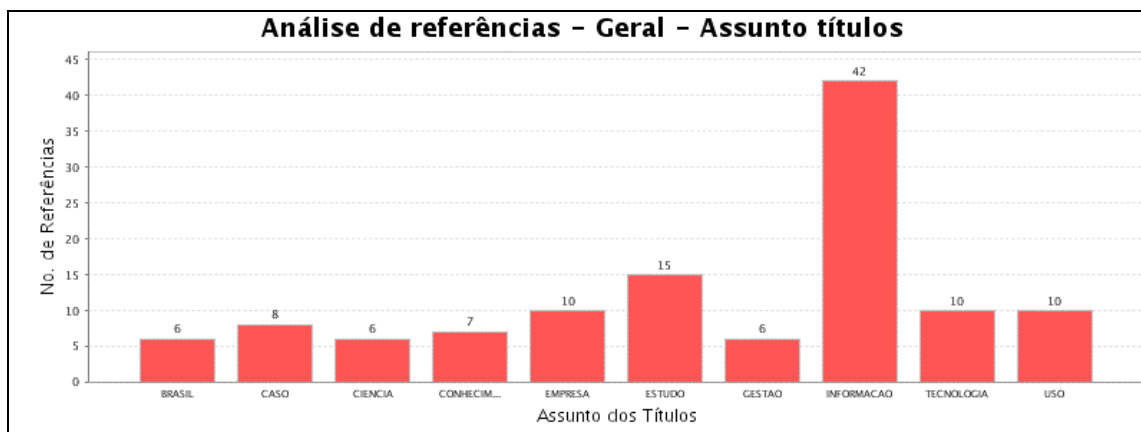


Gráfico 6 – Análise Geral das citações por Autor



**Gráfico 7 – Análise Geral das citações por Assunto dos Títulos**

## **4.2 Resultados por Linha de Pesquisa**

Após a Análise Geral, foi possível extrair um “retrato” mais profundo do PPGCI, através das especificidades de cada Linha de Pesquisa. Como o PPGCI da UFMG possui três Linhas de Pesquisa, os resultados foram coletados para cada uma delas, vislumbrados nas subseções abaixo.

### **4.2.1 *Gestão da Informação e do Conhecimento - GIC***

Para a Linha de Pesquisa Gestão da Informação e do Conhecimento (GIC), com 30 documentos e média aritmética de 80 referências por documento, foram obtidos os seguintes resultados:



Gráfico 8 – Análise das citações por Ano para a Linha de Pesquisa GIC

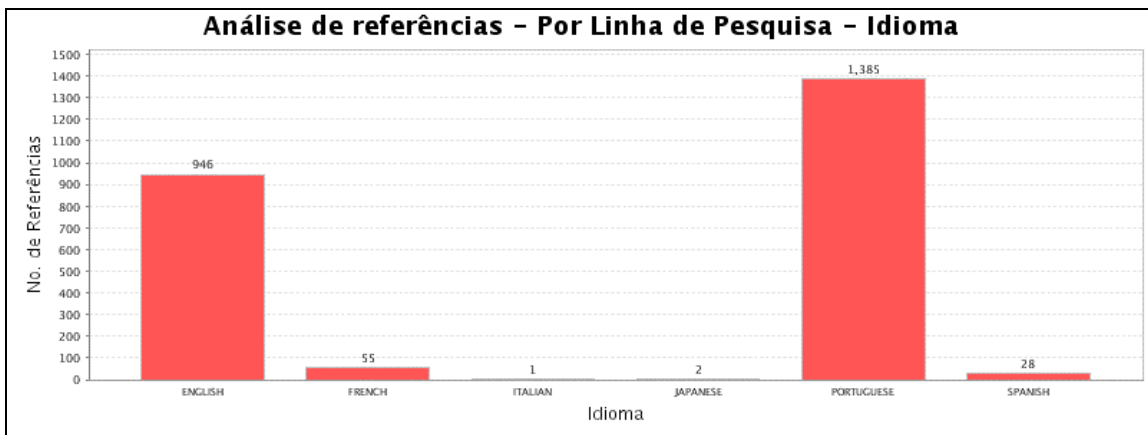


Gráfico 9 – Análise das citações por Idioma para a Linha de Pesquisa GIC



Gráfico 10 – Análise das citações por País de Publicação para a Linha de Pesquisa GIC



Gráfico 11 – Análise das citações por Tipo de Publicação para a Linha de Pesquisa GIC

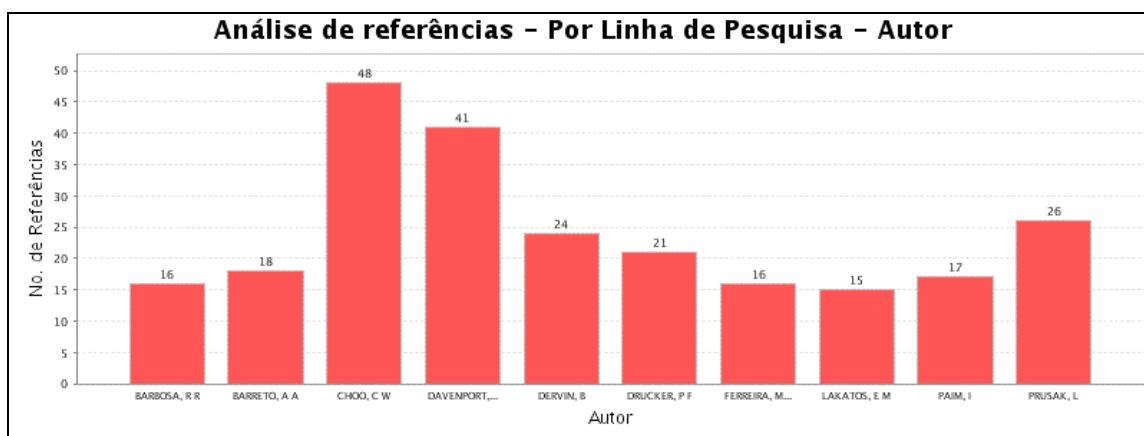


Gráfico 12 – Análise das citações por Autor para a Linha de Pesquisa GIC

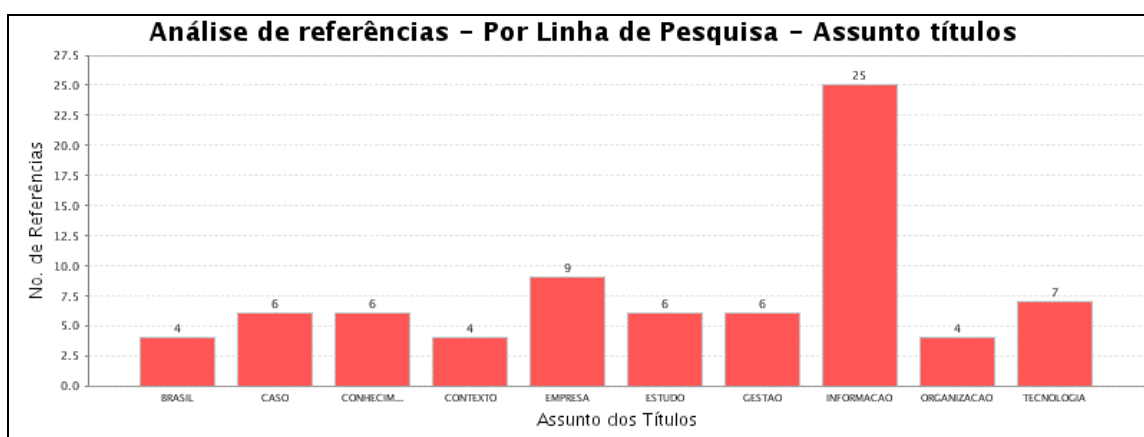


Gráfico 13 – Análise das citações por Assunto dos Títulos para a Linha de Pesquisa GIC



Gráfico 14 – Análise das citações por Assunto das Referências para a Linha de Pesquisa GIC

#### 4.2.2 *Informação, Cultura e Sociedade - ICS*

Para a Linha de Pesquisa Informação, Cultura e Sociedade (ICS), com 18 documentos e média aritmética de 75 referências por documento, foram obtidos os seguintes resultados:

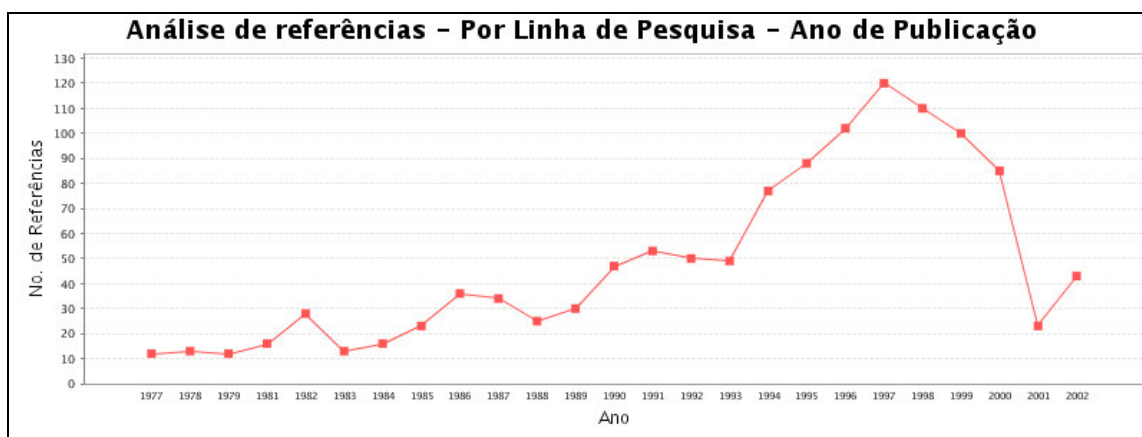


Gráfico 15 – Análise das citações por Ano para a Linha de Pesquisa ICS



Gráfico 16 – Análise das citações por Idioma para a Linha de Pesquisa ICS

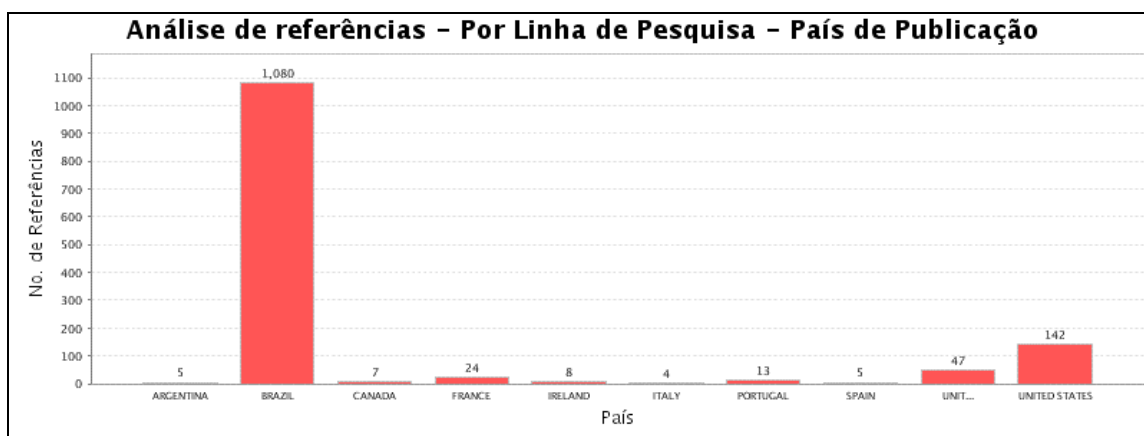


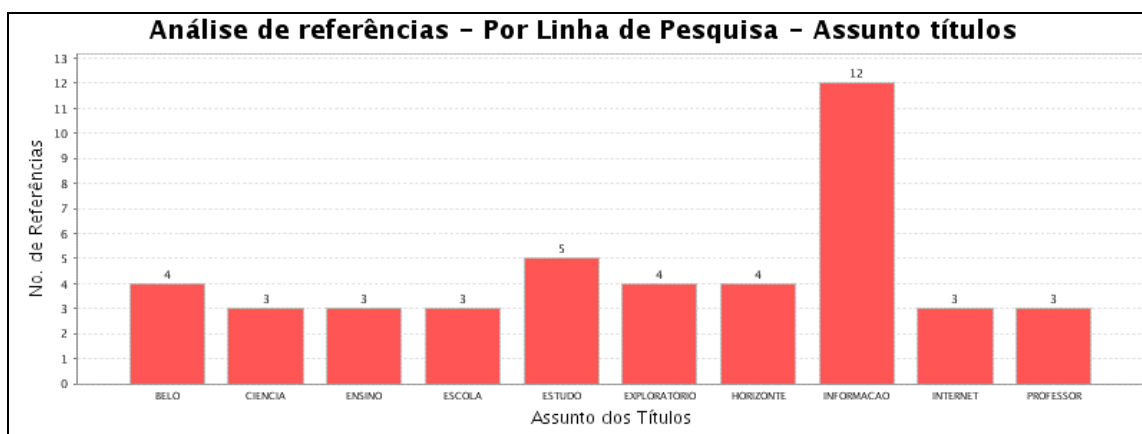
Gráfico 17 – Análise das citações por País de Publicação para a Linha de Pesquisa ICS



Gráfico 18 – Análise das citações por Tipo de Publicação para a Linha de Pesquisa ICS



**Gráfico 19 – Análise das citações por Autor para a Linha de Pesquisa ICS**



**Gráfico 20 – Análise das citações por Assunto dos Títulos para a Linha de Pesquisa ICS**



**Gráfico 21 – Análise das citações por Assunto das referências para a Linha de Pesquisa ICS**



### 4.2.3 Organização e Tratamento da Informação - OTI

Para a Linha de Pesquisa Organização e Tratamento da Informação (OTI), com apenas 7 documentos e a média aritmética de 103 referências por documento, foram obtidos os seguintes resultados:

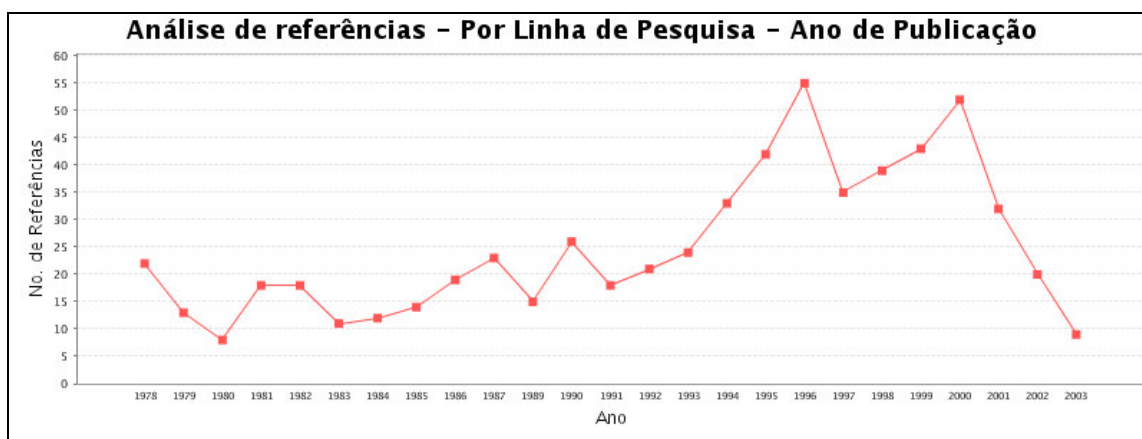


Gráfico 22 – Análise das citações por Ano para a Linha de Pesquisa OTI



Gráfico 23 – Análise das citações por Idioma para a Linha de Pesquisa OTI

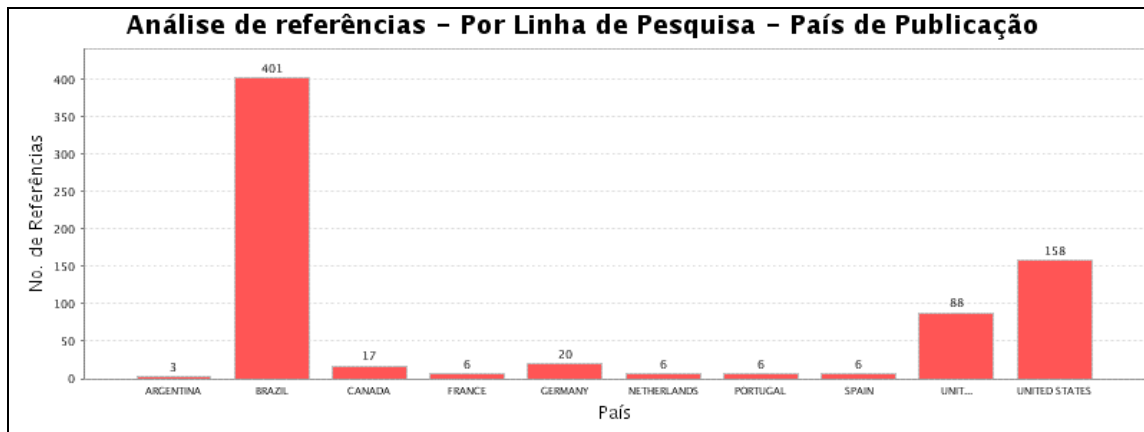


Gráfico 24 – Análise das citações por País de Publicação para a Linha de Pesquisa OTI



Gráfico 25 – Análise das citações por Tipo de Publicação para a Linha de Pesquisa OTI

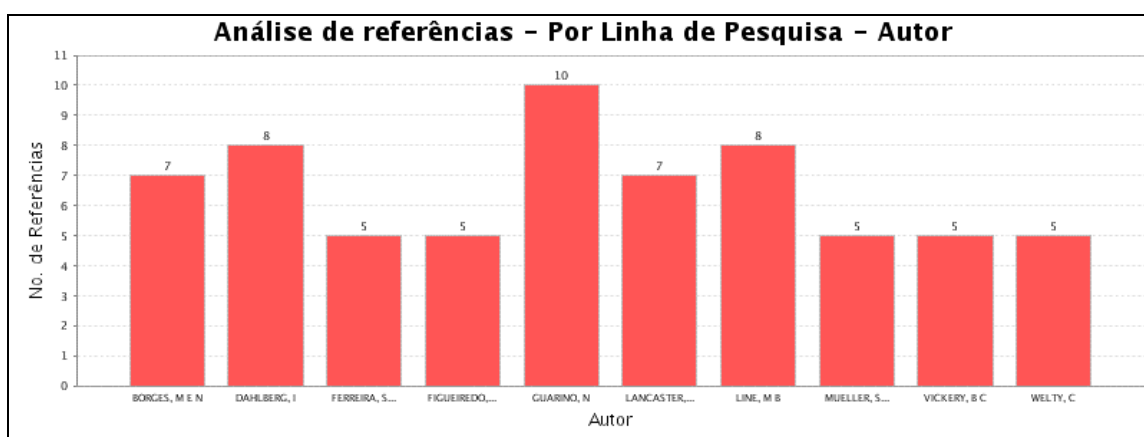
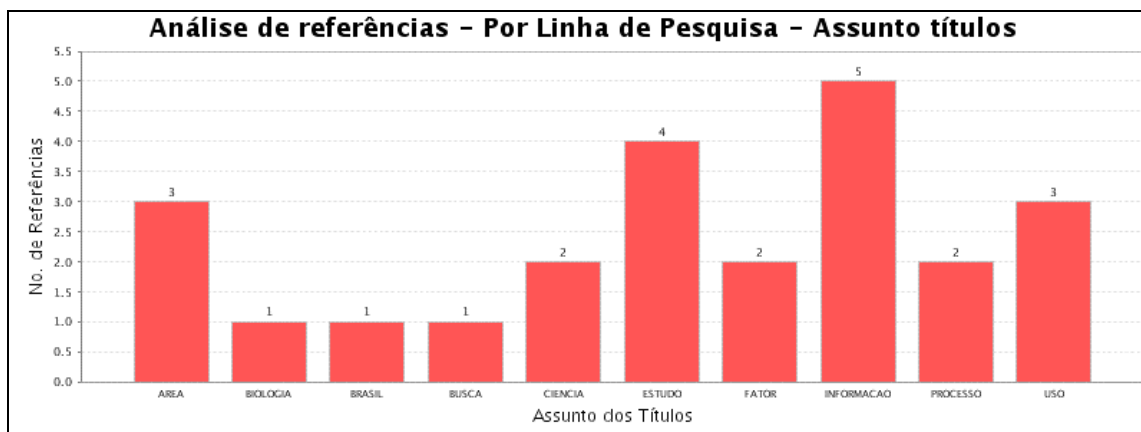
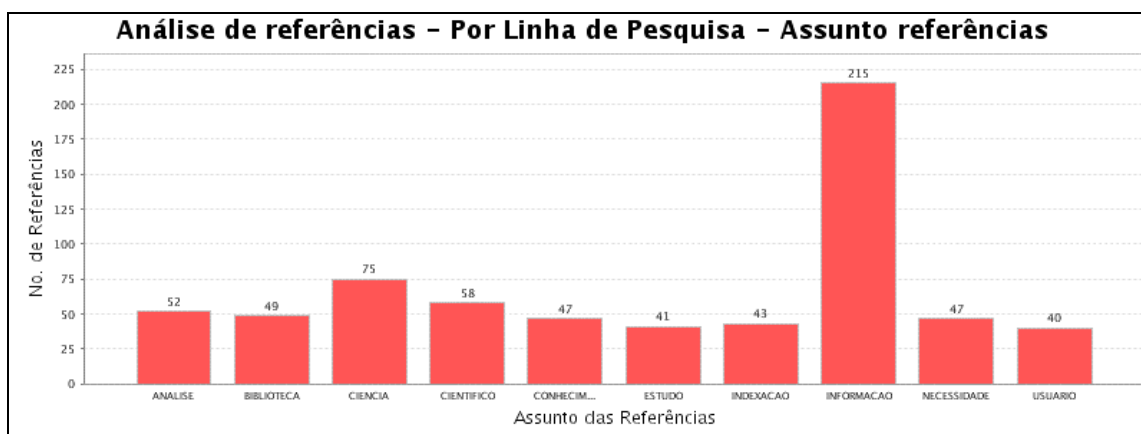


Gráfico 26 – Análise das citações por Autor para a Linha de Pesquisa OTI



**Gráfico 27 – Análise das citações por Assunto dos Títulos para a Linha de Pesquisa OTI**



**Gráfico 28 – Análise das citações por Assunto das Referências para a Linha de Pesquisa OTI**

### 4.3 Resultados por Orientador

A terceira variável independente baseia-se na análise das citações por orientador, com a idéia de que o PPGCI, ou qualquer outro centro de pesquisa científica, deve ter um acompanhamento dos seus membros, uma vez que um dos papéis de qualquer orientador é o de indicar referências aos seus orientados.

Para a apresentação dos resultados encontrados com essa variável, foi escolhido o orientador com maior número de orientados entre os documentos presentes na biblioteca digital trabalhada. O orientador em questão possui 7 orientados, resultando na média aritmética de 70 referências por documentos e os resultados são apresentados abaixo.

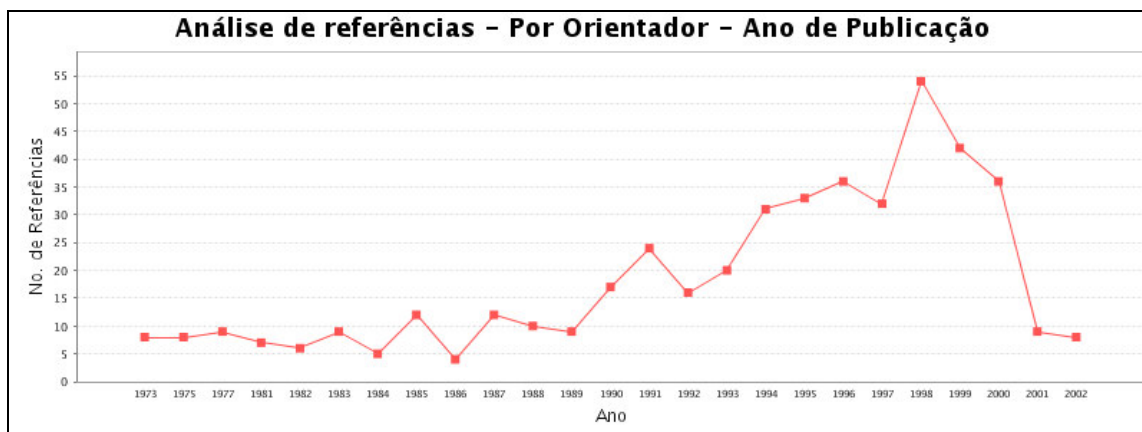


Gráfico 29 – Análise das citações por Ano para o Orientador selecionado

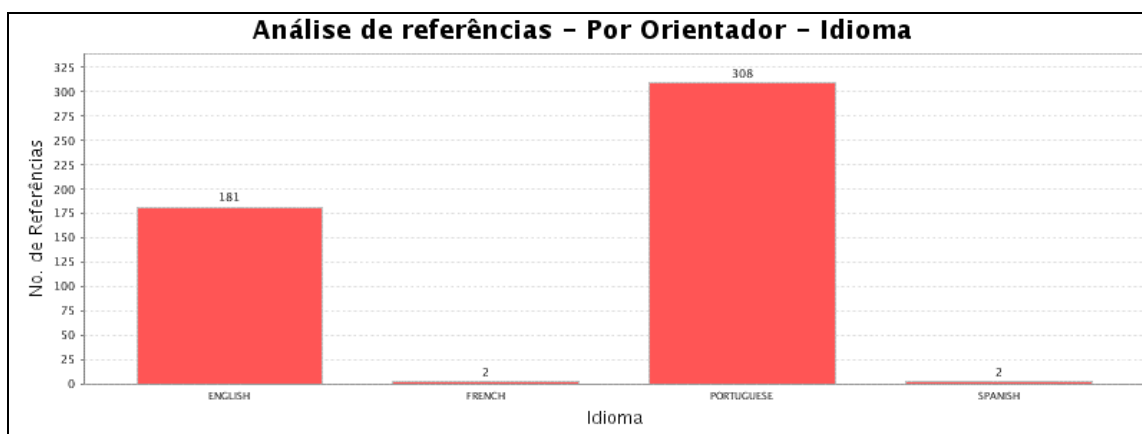


Gráfico 30 – Análise das citações por Idioma para o Orientador selecionado

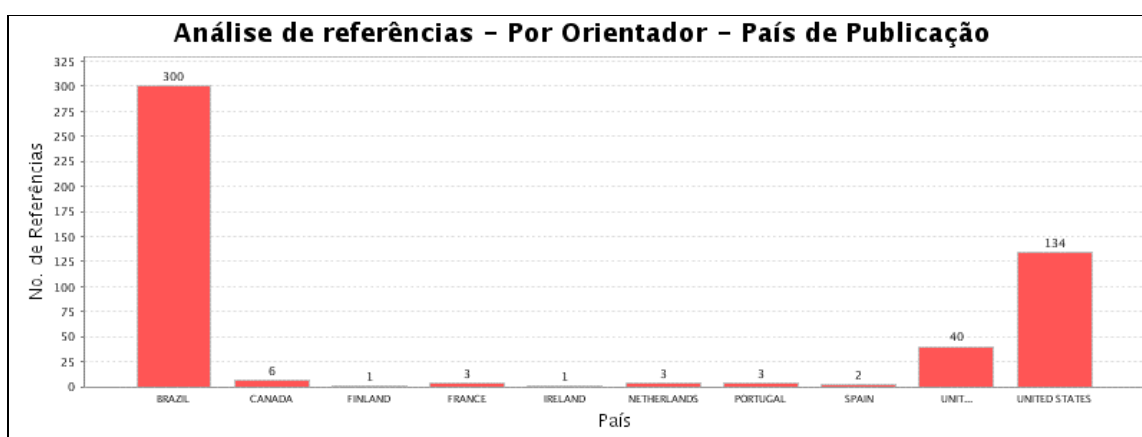


Gráfico 31 – Análise das citações por País de Publicação para o Orientador selecionado



Gráfico 32 – Análise das citações por Tipo de Publicação para o Orientador selecionado

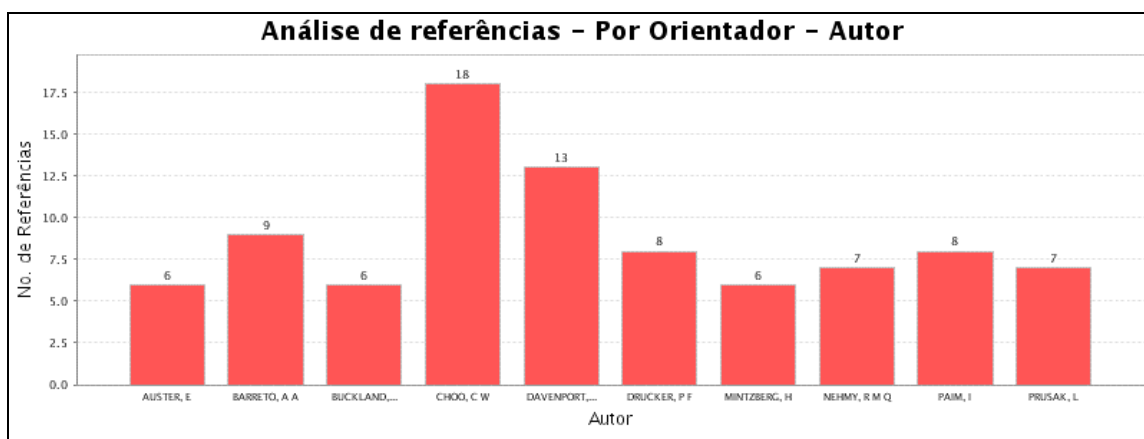


Gráfico 33 – Análise das citações por Autor para o Orientador selecionado

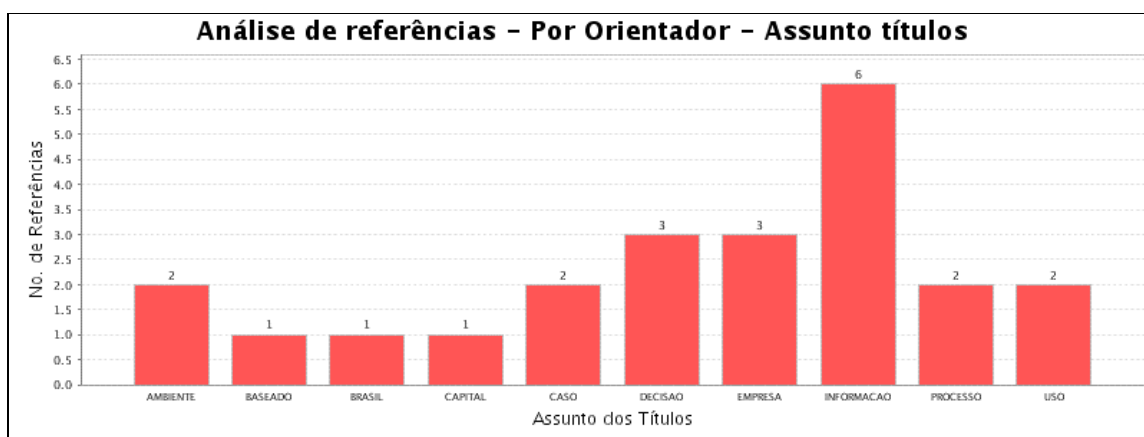
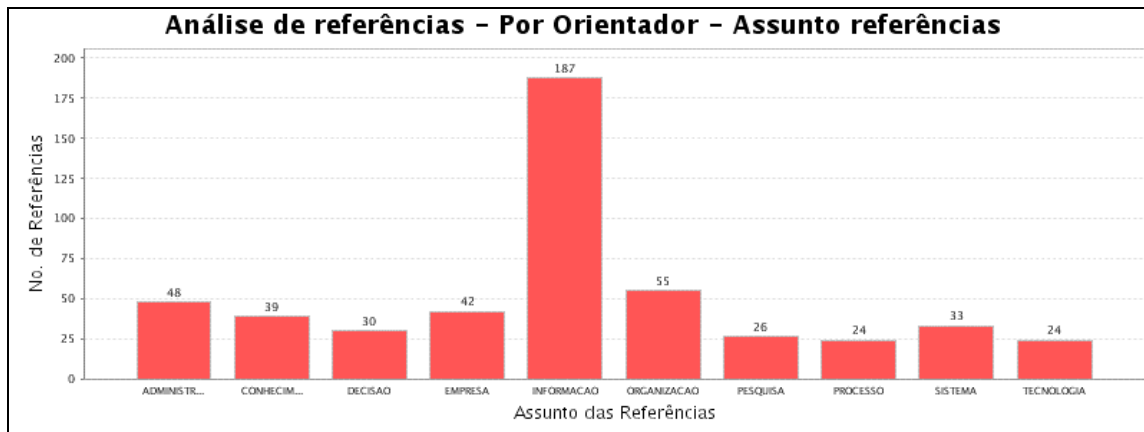


Gráfico 34 – Análise das citações por Assunto dos Títulos para o Orientador selecionado



**Gráfico 35 – Análise das citações por Assunto das Referências para o Orientador selecionado**

## **5 *Análise dos resultados***

Com o capítulo anterior, pudemos vislumbrar o aspecto quantitativo da presente pesquisa, uma vez que foram realizadas atividades com a finalidade de mensurar variáveis pré-estabelecidas, verificar e explicar sua influência sobre outras variáveis, mediante a análise do número de incidências. Na pesquisa quantitativa, o pesquisador descreve, explica e prediz (CHIZZOTTI, 1998).

Já, no presente capítulo, a intenção é apresentar uma análise qualitativa dos dados coletados. Essa etapa se fundamenta nas análises realizadas a partir da significação que os dados coletados dão aos atos. Nesse tipo de análise, o pesquisador participa, compreende e interpreta (CHIZZOTTI, 1998).

Uma vez que a análise qualitativa não foi o objetivo principal deste trabalho, nas seções desse capítulo são apresentadas inferências não tão aprofundadas, de modo que estas últimas devem ser contempladas em trabalhos futuros, dando seqüência a este.

Todas as variáveis dependentes podem ser usadas pelos coordenadores de um centro de pesquisas, a fim de ajudá-los na melhoria ou correção do “retrato” dos trabalhos resultantes de sua produção científica.

Avaliações sobre os anos das publicações podem representar se os trabalhos têm acompanhado a produção mais recente ou se têm privilegiado o conhecimento mais antigo. Analisando os idiomas e os países das publicações a coordenação toma conhecimento daqueles que têm privilegiado mais o conhecimento nacional ou internacional. De posse dos tipos de publicação mais utilizados, descobrirão quais têm sido aqueles trabalhos que têm se esforçado a acompanhar a produção mais recente, citando um maior número de trabalhos de Congressos e artigos de periódicos e que têm contribuído para a continuidade dos trabalhos científicos, com citações para Teses e Dissertações.

Ao analisar os autores mais citados nos trabalhos, pode-se compará-los com os outros trabalhos de uma mesma Linha de Pesquisa, o que pode alertar para uma possível falta de coesão do corpo

docente ou até pior, uma falta de trabalho coletivo, apresentando resultados de pesquisas isoladas, fora de grupos de pesquisa, quando cada um utiliza fontes diferentes do conhecimento.

Análises sobre os assuntos presentes nos títulos dos trabalhos e nos títulos das referências utilizadas por esses podem indicar o uso de termos que não são comuns em uma área de pesquisa, ou de uma determinada Linha de Pesquisa. Mais importante ainda, com o acompanhamento dos conceitos usados no decorrer do tempo, podemos verificar quando um assunto começou a ser trabalhado pelos pesquisadores e o quanto ele ainda tem sido pesquisado pelo centro de pesquisa; no caso deste trabalho, pelo PPGCI.

## **5.1 Análise Geral**

A primeira análise a ser feita, como na parte quantitativa, diz respeito à Análise Geral. A partir desta análise, tenta-se retirar algumas relações causais dos fatos.

Ao analisar a primeira variável dependente, sobre o ano de publicação das citações feitas, vê-se que há picos de citações nos anos de 1982, com cerca de 70 citações; 1986, com cerca de 100; 1991, perto de 180; 1994 com quase 250; e o auge em 1998, com mais de 400 citações feitas. Sendo assim, percebe-se que os picos ocorrem de 4 em 4 anos, quando deve ocorrer renovação no referencial teórico ou novos documentos são publicados e passam a ser mais conhecidos, valorizados e, com isso, mais citados.

Analisando a segunda variável dependente verificamos a presença maciça da língua portuguesa, com quase 65% do total das referências. Em segundo lugar, como era imaginado, aparece o Inglês, aparecendo em 32% das 4.495 referências. Outros idiomas estão presentes, como o Francês, Espanhol, Italiano, Japonês e Alemão, em ordem decrescente, completando os 3% restantes.

Sobre os países de publicação, o resultado encontrado acompanha a distribuição dos idiomas, pois o Brasil é a origem das referências em 64% e os países de língua inglesa, como os Estados Unidos, Reino Unido e Canadá, ocupam 30%. Os resultados dessas duas variáveis demonstram uma grande valorização da produção científica nacional, provavelmente decorrente da falta de acesso à produção internacional ou da tentativa de se privilegiar a produção nacional, o que demandaria análises mais profundas.



A análise dos tipos de publicação remete a conclusões interessantes. Através dela, podemos verificar se a produção do PPGCI acompanha o conhecimento mais atualizado, com citações para artigos científicos e trabalhos apresentados em congressos, ou se são utilizadas fontes de conhecimento já consolidado, como os livros. A segunda alternativa é percebida, com 49% das citações para livros e 39% para artigos de periódicos e trabalhos de congressos, o que pode não ser muito interessante, pois o conhecimento do estado da arte pode não estar sendo referenciado. Outra conclusão que pode ser inferida diz respeito ao pequeno número de referências para Dissertações e Teses, de apenas 167 ou 4%, o que significa uma pequena coesão entre as pesquisas do PPGCI, uma vez que o que já foi produzido no PPGCI não foi muito referenciado pelos trabalhos analisados, bem como os trabalhos produzidos por outros cursos de Pós-Graduação do Brasil, o que pode sugerir certa falta de continuidade das pesquisas e trabalho muito individualizado, no estilo cada um por si.

A quinta variável dependente consiste na autoria das referências utilizadas. Através da Análise Geral, foi identificado o autor *CHOO, C W* como o mais citado, com 55 citações ou 1,22%; seguido por *DAVENPORT, T*, com 43 citações; *BARRETO, A A*, com 31; e *LEVY, P*, com 30 citações. Ter um pesquisador nacional como o terceiro mais citado entre as teses e dissertações do PPGCI é um bom indicativo, pois demonstra que se tem valorizado a produção nacional. Os autores mais citados foram os mesmos da Linha de Pesquisa GIC, possivelmente por representarem o maior número de documentos dessa linha, 30 documentos.

A sexta e última variável dependente trabalhada na Análise Geral são os assuntos extraídos dos títulos dos documentos analisados. Como são trabalhos da área da Ciência da Informação, o termo *INFORMAÇÃO* foi o mais utilizado, com 42 dentre os 280 termos, representando 15%. Em segundo lugar está o termo *ESTUDO*, com 5,35%, seguido de *EMPRESA*, *TECNOLOGIA* e *USO* com 3,57% cada um deles. Como ocorrido com os autores, os assuntos também sofreram maior impacto pelos documentos da Linha de Pesquisa GIC, por apresentarem 54% do universo de documentos analisados os termos mais utilizados obtidos na Análise Geral foram semelhantes aos da Análise sobre a Linha GIC.

## 5.2 Análise por Linha de Pesquisa

A análise das citações dos documentos da biblioteca digital quando subdivididos em Linhas de Pesquisa podem representar especificidades de cada uma delas, o que não poderia ser identificado na análise geral. O Programa de Pós-Graduação em Ciência da Informação da UFMG se divide em três Linhas de Pesquisa: Gestão da Informação e do Conhecimento (GIC), Informação, Cultura e Sociedade (ICS) e Organização e Tratamento da Informação (OTI). Nos parágrafos seguintes são detalhadas as conclusões extraídas das diferenças entre as três Linhas.

No que diz respeito aos anos das publicações utilizadas nas referências, para a GIC, podemos perceber que os picos ocorrem, inicialmente, de 4 em 4 anos, mas, a partir de 1991, esse espaço é menor, caindo para 2 anos. Como se trata de uma Linha de Pesquisa voltada a assuntos que estão mais em voga, temas da atualidade, as citações devem acompanhar essa evolução buscando os trabalhos mais recentes, o que pôde ser observado com os documentos analisados. Já, para a Linha ICS, a absorção dos novos conhecimentos é mais lenta, o que é demonstrado com picos de 4 a 6 anos. Com a Linha OTI, a distribuição das citações é mais homogênea, pois o maior pico do gráfico não ocorre com um número de citações maior do que o dobro dos outros picos, fato constatado nas outras duas Linhas, na GIC com diferença de até 200% e na ICS de 150%.

Tomando como exemplo o ano de 1991, o penúltimo pico do gráfico nas três Linhas, percebe-se que na Linha GIC o número de citações representa apenas 1/3 do número encontrado no ano de 1998. Com a Linha ICS ocorre diferente, o número de citações para o ano de 1991 representa quase a metade de 1998. Isso significa que a Linha ICS possui uma distribuição das citações mais uniforme, fazendo referência a trabalhos mais recentes, sem deixar de lado os mais antigos, uma vez que análises sociais sempre são feitas com o olhar no passado, na tentativa de explicar o presente. Com a Linha OTI essa distribuição é mais uniforme ainda, pois possui vários picos (anos de 1978, 1981, 1987, 1990) que possuem metade das citações encontradas no maior pico, no ano de 1996.

Sobre o idioma das referências, é evidente o uso intensivo da língua portuguesa e, em segundo lugar, da língua inglesa. O diferencial é o porcentual de cada uma delas nas citações feitas pelos documentos contidos na Linha analisada. Na Linha GIC o Português representou 57% e o Inglês 39%; na ICS o Português apareceu em 80% das citações e o Inglês em apenas 16%; na terceira Linha, a OTI, o Português esteve presente em 57% e o Inglês em 39%. Com essa análise,

percebemos que as especificidades nacionais são seriamente levadas em consideração nos trabalhos da Linha ICS.

Seguindo os resultados encontrados para o idioma, analisando o país de publicação das referências obtemos as mesmas distribuições entre o Brasil e Estados Unidos. Vale ressaltar que os documentos da Linha de Pesquisa OTI são os únicos a referenciar trabalhos da Alemanha, e ainda com um número razoável, sendo o quarto país mais citado. Para a Linha GIC, os trabalhos de origem francesa são muito utilizados, sendo, também, o quarto país em ordem decrescente.

A quarta variável dependente analisada consiste na análise da tipologia das referências. Nas três linhas, os tipos mais utilizados foram livros, artigos de periódicos e trabalhos apresentados em congressos. Seguindo as conclusões apresentadas na análise geral, percebe-se que a Linha OTI é a única a apresentar mais citações para artigos do que para livros, o que pode considerar esta Linha como a que mais se preocupa com a atualização do conhecimento, utilizando referências, possivelmente, de ponta. Por outro lado, um aspecto negativo foi encontrado na Linha GIC, uma vez que apenas 2,85% das citações foram para Teses e Dissertações, demonstrando a falta de continuidade entre os trabalhos dessa Linha, percentual este de 4,29% na Linha ICS e 4,9% na Linha OTI. É interessante perceber que os trabalhos apresentados em Congressos não foram muito privilegiados em nenhuma das Linhas, uma vez que representam o conhecimento mais atual, mais ainda do que os artigos de periódicos.

A análise dos autores mais citados é interessante por revelar quais têm sido aqueles que mais têm exercido influência sobre os trabalhos produzidos pelos alunos do PPGCI. Aquilo que fora imaginado aconteceu, uma vez que não houve interseção entre os conjuntos dos autores mais citados das três Linhas, evidenciando os autores específicos de cada área. Na Linha GIC, os autores mais citados, em ordem decrescente, foram *CHOO, C W*; *DAVENPORT, T*; *PRUSAK, L* e *DERVIN, B*. Para a Linha de Pesquisa ICS, as bases do conhecimento foram extraídas dos trabalhos de *MARTELETO, R*; *GOMEZ, M N G*; *GIDDENS, A* e *CARDOSO, A M P*, do mais para o menos citado. Já, para a Linha OTI, os autores mais citados foram *GUARINO, N*; *LINE, M B*; *DAHLBERG, I* e *LANCASTER, F W*.

O assunto mais utilizado nos títulos dos trabalhos é *INFORMAÇÃO*, para todas as três Linhas. Os termos em segundo, terceiro e quarto lugares representam assuntos específicos de cada área de estudo, onde para a GIC temos *EMPRESA, TECNOLOGIA* e *CONHECIMENTO*; para a ICS

temos ESTUDO, BELO HORIZONTE e EXPLORATÓRIO; e para a Linha OTI estão ESTUDO, ÁREA e USO.

A última variável dependente analisada diz respeito aos títulos das referências utilizadas. O assunto com maior número de ocorrências também foi o termo INFORMAÇÃO e em todas as Linhas esse termo apresentou uma grande distância para os segundos termos mais utilizados. Com termos remetendo a conceitos semelhantes aos termos mais utilizados nos títulos dos trabalhos do PPGCI, encontramos para os títulos das referências da Linha GIC, TECNOLOGIA, ADMINISTRAÇÃO e EMPRESA, na Linha ICS os termos CIÊNCIA, SOCIAL e BRASIL e, por fim, na Linha OTI foram encontrados os termos CIÊNCIA, CIENTÍFICO e ANÁLISE.

### **5.3 Análise por Orientador**

A idéia de criar a variável independente de orientador originou-se da curiosidade de poder comparar o resultado entre as referências utilizadas pelos orientados dos orientadores, proporcionando à coordenação do centro de pesquisa uma forma eficaz de avaliação de seus membros.

Para os trabalhos defendidos pelos orientados do orientador selecionado, percebemos que os anos das referências alternam seus picos de 2 em 2 anos, para referências mais antigas, de 1973 a 1987, e de 4 em 4 anos de 1987 a 1997. Como ocorreu com a análise geral e a análise por Linha de Pesquisa, o ano mais citado foi o de 1998.

Os idiomas mais utilizados nas referências foram o Português e o Inglês, com cerca de 62 e 37%, respectivamente, no universo de referências dos trabalhos dos orientados. Dessa mesma forma, o Brasil e os países de língua inglesa, como o Canadá, Reino Unido e Estados Unidos, foram as origens mais utilizadas das publicações, estando presentes em 61 e 37%, respectivamente.

A análise dos tipos de publicação mais utilizados leva a concluir que o orientador e seus orientados não têm se preocupado com a produção recente, fazendo uso de livros em 59% dos casos. Outra constatação é a falta de continuidade das pesquisas, utilizando apenas 3% das referências para Teses e Dissertações.

Ao verificar os autores mais citados, percebemos que o orientador deve fazer parte da Linha de Pesquisa GIC, pois foram encontrados os nomes de *CHOO, C W; DAVENPORT, T* e *BARRETO,*

A A como os mais citados, em ordem decrescente. Se o professor não fizer parte dessa Linha, conclui-se que este deveria ser, uma vez que os autores considerados como os mais importantes da Linha estão contemplados por ele.

Os assuntos extraídos tanto dos títulos dos trabalhos quanto dos títulos das referências também demonstram qual a Linha de Pesquisa que o orientador selecionado deve fazer parte. Foram utilizados com maior frequência os termos INFORMAÇÃO, EMPRESA, DECISÃO e PROCESSO nos títulos dos trabalhos e INFORMAÇÃO, ORGANIZAÇÃO, ADMINISTRAÇÃO e EMPRESA nos títulos das referências feitas pelos trabalhos analisados.

## **6** *Considerações Finais*

A bibliometria, um assunto discutido desde a década de 60 (SENGUPTA, 1992), atualmente tem sido abordada cada vez mais nos trabalhos científicos. A análise feita por Maia e Caregnato (2004) associa a nova fase da bibliometria com a situação tecnológica atual, onde as Tecnologias da Informação e Comunicação estão mais evoluídas e disponíveis, proporcionadas pela Internet e a forte relação destas com aspectos relacionados a bibliotecas digitais. Essa nova situação tecnológica é observada principalmente nos países subdesenvolvidos, não tão valorizados cientificamente, como o Brasil, onde o acesso à informação técnico-científica tem estado cada dia mais disponível.

Dado a esse crescente volume informacional produzido mundialmente (LE COADIC, 1996), para encontrar informações relevantes, é necessária a utilização de ferramentas que auxiliem na sua recuperação. Visando criar um mecanismo de tratamento da informação que alcançasse esse objetivo, este trabalho apresentou uma ferramenta que busca colaborar com a divulgação e disseminação dos produtores do conhecimento científico nacional - o índice de citação. A partir deste índice, foram implementadas formas de análises bibliométricas a fim de subsidiar seus usuários no que diz respeito ao acompanhamento sistemático de sua produção científica.

Analisando o tempo despendido com a parte experimental do trabalho, percebemos que o mesmo foi elevado, cerca de 2 meses, dado o grande número de documentos a trabalhar. Esse dispêndio não deve ocorrer com o usuário do sistema no seu dia a dia, pois novas TD's são defendidas e aprovadas com um espaço de tempo entre elas, após um necessário treinamento deste com a nova ferramenta apresentada. Como a biblioteca digital utilizada no desenvolvimento deste trabalho já contempla um grande número de documentos, justifica-se seu uso, principalmente para o Programa de Pós-Graduação da UFMG. Após a entrega do sistema à coordenação do Programa, ao utilizá-lo de forma constante e consistente, permitirá o acompanhamento com a inclusão contínua de novas teses e dissertações.

Devido aos problemas encontrados pela falta de padronização dos documentos analisados durante este trabalho, apresentamos à coordenação do PPGCI um possível padrão de apresentação dos documentos entregues à secretaria do Programa. Esse padrão não foi apresentado com o objetivo

de dar uma resposta final e absoluta a todos os problemas, mas para alertá-la da dificuldade de análise gerada pela baixa padronização dos documentos, apontando possíveis soluções. Os problemas encontrados durante o trabalho, identificados e sanados a tempo, criam empecilhos, podendo levar a análises errôneas, caso muitas referências bibliográficas fossem redigidas de forma incorreta.

Com relação aos resultados encontrados durante a etapa experimental do trabalho, onde as teses e dissertações do PPGCI foram analisadas, é difícil discernir se as análises das variáveis dependentes apresentaram resultados positivos ou negativos, o que depende da estratégia de cada centro de pesquisa. Dependendo do que é almejado pela coordenação, deve-se acompanhar esses indicadores a fim de melhorar o “retrato” da produção científica do centro de pesquisa, indicando possíveis alterações nas referências utilizadas, como, por exemplo, nas indicações das referências do orientador aos seus orientados.

A ferramenta desenvolvida se mostrou eficiente nos objetivos traçados, sobretudo no que diz respeito à busca do autoconhecimento de um centro de pesquisa, relativo à sua produção científica. De posse do material gráfico produzido e de outros não explicitados neste trabalho, todos os envolvidos com pesquisa podem tirar vantagens com o uso desse aplicativo. Os alunos têm a possibilidade de acessar referências utilizadas por colegas da mesma área, facilitando seu levantamento bibliográfico, os professores podem acompanhar a sua evolução com o trabalho realizado junto aos seus orientados e a coordenação do Programa de Pós-Graduação tem em mãos um mecanismo de acompanhamento de todo seu universo de trabalho, com a possibilidade de melhorar o “retrato” tirado a cada ano, por exemplo.

É interessante ressaltar que a ferramenta não é atrelada a uma área específica do conhecimento, viabilizando um fornecedor de descobertas para qualquer centro de pesquisa, pois não se vincula, apenas, à área da Ciência da Informação. Com isso, seria de extrema importância um projeto em grande escala que implementasse a ampliação da utilização de ferramentas como a apresentada neste trabalho, principalmente nos grandes centros de pesquisa nacionais, como as universidades públicas. De posse de informações precisas sobre a produção científica de cada universidade, comparativos podem ser realizados, de modo a utilizar a análise bibliométrica como fator de qualificação das universidades junto à CAPES.

Essa utilização em massa teria outro aspecto positivo, no que diz respeito à divulgação do material científico à toda comunidade, corroborando com as idéias dos autores Pacheco e Kern (2001), como um retorno de seus investimentos, mesmo que indiretamente. Vê-se que essa situação não é de tudo uma utopia, uma vez que, como analisado durante esta pesquisa, a grande maioria dessas universidades já utilizam mecanismos próprios de disponibilização de suas teses e dissertações, faltando unificá-los, para que alcancem os três princípios almejados por Soergel (2002): praticidade, novo meio para o trabalho intelectual e suporte à colaboração e às comunidades na prática. Essa unificação pode aumentar a visibilidade de seus pesquisadores, podendo melhorar o número de citações feitas para as teses e dissertações brasileiras, como afirma Lawrence (2001).

Certos desse sonho, de trabalharmos juntos, em sintonia com outros pesquisadores de outras regiões do Brasil, seguimos em frente com a busca incessante de meios a facilitar nossa comunicação científica, o “coração da ciência” (MEADOWS, 1999) e, mais do que tudo, legitimamente nacional.



## **7 *Trabalhos Futuros***

Como todo trabalho científico, a pesquisa relatada nesta dissertação não contemplou todas as possíveis facetas da área em que atuou, principalmente no que diz respeito às citações, assunto há muito discutido, porém nunca esgotado. Seja pela falta de tempo ou de conhecimento necessário, várias alternativas de melhoria e aprofundamento, a seguir relatadas, não foram abordadas.

Em ambos os momentos da pesquisa, etapa aplicada e etapa experimental, várias frentes de pesquisa poderiam ser criadas a fim de mergulhar mais profundamente no trabalho. Porém, antes de entrar nas fases da implementação, cabe ressaltar a necessidade de se criar um padrão para as Teses e Dissertações do PPGCI que seja seguido à risca pelos alunos do Programa. Em vários momentos do trabalho a falta de padrão e até mesmo a incidência de erros, principalmente no que diz respeito à formatação das referências bibliográficas, demandaram um tempo maior do que o previsto, até que o determinado problema fosse sanado.

Na fase de implementação, onde o aspecto tecnológico prevalece, em vários pontos as TIC's podem colaborar com o aperfeiçoamento do trabalho, automatizando etapas realizadas pelo usuário.

A etapa de adição do documento à biblioteca digital, realizada pelo usuário administrador do sistema, poderia ser alterada de forma a criar um mecanismo inteligente de extração dos metadados do padrão Dublin Core. Ao dar entrada no sistema de uma nova tese ou dissertação (TD) no formato PDF, a ferramenta extrairia automaticamente essas informações, aumentando muito a produtividade de quem fosse utilizar o sistema, principalmente no que tange às referências bibliográficas, uma vez que todas as 4.495 referências foram adicionadas uma a uma pelo usuário.

A etapa seguinte, extração das variáveis dependentes de cada referência bibliográfica, também é passível de automatização ou semi-automatização. Cabe a futuros trabalhos apresentar um mecanismo onde o ano da publicação, seu tipo, a autoria, o título, o idioma utilizado e o país de origem sejam previamente extraídos de cada referência, restando ao usuário apenas a correção de possíveis erros dos autores das TD's ou da extração automática. Já existem muitos trabalhos

nesse intuito, como os publicados pelo instituto de pesquisa que desenvolve o Citeseer, NEC<sup>15</sup> e Universidade de Pensilvânia<sup>16</sup> (LAWRENCE, 2001; LAWRENCE; GILES; BOLACKER, 1999).

Agindo também em busca da otimização do trabalho realizado pelo usuário, pode-se incluir nesse desdobramento da ferramenta a questão da criação dos arquivos de autoridade, tanto de autores, quanto de assuntos. Para os autores, podem ser trabalhados mecanismos de consulta por trabalhos científicos como o *Google Scholar*<sup>17</sup>, *Web of Science*<sup>18</sup>, *ACM Portal*<sup>19</sup>, a fim de tentar criar interseção de trabalhos entre as variações dos nomes dos autores, indicando ser as mesmas pessoas. Já, com relação aos arquivos de autoridades de assuntos, podem ser implementados agentes inteligentes de *software* que utilizem dicionários eletrônicos a fim de conectar as variações de termos, como masculino e feminino, singular e plural e mais ainda, com dicionários de línguas, como de Inglês para o Português, ou vice-versa, retirando as ambigüidades até mesmo entre línguas distintas de forma automática. Da mesma forma como indicada para as variáveis dependentes no parágrafo anterior, dessa nova maneira o usuário do sistema teria, apenas, a tarefa de verificar se os arquivos de autoridade gerados estão de acordo com a realidade.

Analisando a parte experimental do trabalho, com as duas partes da pesquisa, quantitativa e qualitativa, também ficam abertas algumas novas possibilidades de estudo. Sobre a parte quantitativa, seria interessante o cruzamento entre algumas variáveis dependentes, a partir da seleção da variável independente, em busca de respostas para outros questionamentos. Podemos citar a quantificação do número de vezes em que um autor é citado no decorrer dos anos para a Linha de Pesquisa GIC, ou a evolução de um determinado assunto no decorrer do tempo, com a possibilidade da descoberta da data em que um assunto começou a ser trabalhado nas TD's do Programa. Pode ser trabalhada, ainda, a inclusão de novas variáveis dependentes, como o periódico dos artigos citados, de modo a descobrir o nível de nossas citações, segundo o conceito

---

<sup>15</sup> Informações sobre o Instituto NEC disponíveis em <http://www.nec-labs.com/>. Acessado em 18/06/2005.

<sup>16</sup> Escola de Ciências da Informação e Tecnologia da Universidade da Pensilvânia mantenedora do Citeseer. Maiores informações em <http://ist.psu.edu/>. Acessado em 18/06/2005.

<sup>17</sup> Google Scholar, acessível no endereço <http://scholar.google.com/>. Acessado em 18/06/2005.

<sup>18</sup> Web of Science, mecanismo de busca de artigos científicos desenvolvido pelo ISI, disponível no endereço <http://isiwebofknowledge.com/>. Acessado em 18/06/2005.

<sup>19</sup> Portal ACM, site disponível no endereço <http://portal.acm.org>. Acessado em 18/06/2005.

dos periódicos na qualificação da CAPES<sup>20</sup>, levando-nos ao conhecimento se são mais citados artigos dos periódicos mais bem avaliados ou não.

De posse de novas análises quantitativas, automaticamente, poderão ser realizadas novas análises qualitativas. Como um desejável trabalho futuro, podemos salientar as importantes descobertas que podem ser feitas, caso a análise qualitativa extrapole aquela feita por este trabalho. Essas novas análises poderiam ser em busca de respostas balizadas e alicerçadas não apenas nos números obtidos, mas também com um conhecimento de causa mais profundo, de modo a explicar a ocorrência dos fatos observados.

---

<sup>20</sup> CAPES, sigla de Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, órgão vinculado ao Ministério da Educação do Brasil. Informações disponíveis em <http://www.capes.gov.br/capes/portal/>. Acessado em 18/06/2005.

## 8 Referências Bibliográficas

1. ADAIR, W. C. Citation Indexes for Scientific Literature? *American Documentation*, 6, 31-32, 1955 apud WEINSTOCK, M. "Citation Indexes", *Encyclopedia of Library and Information Science*, volume 5, pages 16-40, Marcel Drekker, Inc., New York, 1971.
2. ALVARENGA, Lídia. A Institucionalização da Pesquisa Educacional no Brasil. Belo Horizonte: UFMG, 1996 (tese de doutorado).
3. ALVARENGA, Lídia. A teoria do conceito revisitada em conexão com ontologias e metadados no contexto das bibliotecas tradicionais e digitais. *DataGramaZero*, v. 2, n. 6, dez/2001.
4. BAX, Marcello Peixoto. Projeto SABiO – Sistema de Agentes para Bibliotecas Digitais. Relatório Técnico Final submetido à FAPEMIG em julho de 2003.
5. BAX, Marcello Peixoto; LEAL, Igor Campos. Especificação de um Índice de Citação para Biblioteca Digital Auto-incremental, In: XII ENCONTRO NACIONAL DE CENTROS DE INFORMAÇÃO E BIBLIOTECAS DA ÁREA DE COMUNICAÇÃO, ENDOCOM. Setembro de 2002, Salvador, Bahia.
6. BENTES PINTO, V.; MOTA, F. R.; QUEIROZ, N. P. A representação do conhecimento através da análise de citações: o caso da UFC. V Encontro Nacional de Pesquisa em Ciência da Informação, Belo Horizonte, 10-14 nov. 2003.
7. BUCKLAND, Michael K. Information as thing. *Journal of the American Society for Information Science*. 42(5):351-360, 1991.
8. BUSH, Vannevar. As we may think. *The Atlantic Monthly*, July 1945.
9. CARDOSO, Ana Maria Pereira. Retomando possibilidades conceituais: uma contribuição à sistematização do campo da informação social. *R. Esc. Biblioteconomia UFMG*, Belo Horizonte, v. 23, n. 2, p. 107-114, jul-dez/94.
10. CHIZZOTTI, Antonio. Pesquisas em ciências humanas e sociais. São Paulo: Cortez, 1998, 164p.
11. CRONIN, Blaise. The citation process. London: Taylor Graham, 1984, 103p.
12. DIAS, Eduardo Wense. Contexto Digital e Tratamento da Informação. *DataGramaZero*, v. 2, n. 5, out/2001.
13. DUBLIN CORE Data Set elements. On-line. Disponível em: <http://dublincore.org/documents/dces/>. Acessado em 30/06/2005.

14. FAST, Karl; LEISE, Fred; STECKEL, Mike. Synonym Rings and Authority Files. On-line: [http://www.boxesandarrows.com/archives/synonym\\_rings\\_and\\_authority\\_files.php](http://www.boxesandarrows.com/archives/synonym_rings_and_authority_files.php). Acessado em 18/06/2005.
15. FERREIRA, Aurélio Buarque de Holanda. Novo Dicionário da Língua Portuguesa. Rio de Janeiro: Nova Fronteira S/A, 2ª ed, 1986.
16. FRANÇA, Junia Lessa et al. Manual de Normalização de Publicações Técnico-Científicas. Belo Horizonte: Editora UFMG, 6ª ed, 2003, 230p.
17. FREITAS, Elizabeth Cassimiro de. Produção científica: concepções, parâmetros avaliativos e configuração no Brasil. *Tópicos Educacionais*, Recife, v. 8, n. 2, p. 25-42, jul/dez 1990.
18. GARFIELD, Eugene. Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science*, 122(3159), p.108-111, July 1955.
19. GARFIELD, Eugene. Citation Indexing: Its Theory and Application in Science, Technology and Humanities. New York: John Willey & Sons, 1979.
20. GARFIELD, Eugene. Citation Indexes for Retrieval and Research Evaluation, In CONSENSUS CONFERENCE ON THE THEORY AND PRACTICE OF RESEARCH ASSESSMENT, outubro de 1996, Capri.
21. GARFIELD, Eugene. From Citation Indexes to Informetrics: Is the Tail Now Wagging the Dog? *Libri*, vol 48, no. 2, pg. 67-80, 1998.
22. HABERMAS, Jurgen. Knowledge and Human Interests. London: Heineman Educational Books Ltd, 1972.
23. HABERMAS, Jurgen. Pensamento pós-metafísico; estudos filosóficos. Rio de Janeiro: Tempo Brasileiro, 1990
24. ISO 999. Information and documentation – Guidelines for the content, organization and presentation of indexes. 1996.
25. LANCASTER, Frederick Wilfrid. Indexação e resumos: teoria e prática. Brasília: Briquet de Lemos, 1993.
26. LAWRENCE, Steve. Online or invisible? *Nature*, v. 411, n. 6837, p. 521, 2001.
27. LAWRENCE, Steve; GILES, C. Lee; BOLLACKER, Kurt. Autonomous Citation Matching. Proceedings of the Third International Conference on Autonomous Agents, Seattle, Washington, May 1-5, ACM Press, New York, NY, 1999.
28. LE COADIC, Yves-François. A Ciência da Informação. Brasília: Briquet de Lemos, 1996.

29. LEAL, Igor Campos. Especificação de um Índice de Citação para Biblioteca Digital Auto-incremental. Belo Horizonte: UFMG, 2002 (monografia).
30. LUHN, H. P. Keyword-in-context index for technical literature (KWIC index). Yorktown Heights, N. Y., IBM Advanced Systems Development Division, 1959 apud LANCASTER, Frederick Wilfrid. Indexação e resumos: teoria e prática. Brasília: Briquet de Lemos, 1993.
31. MACROBERTS, Michael H.; MACROBERTS, Barbara R. Problems of citation analysis: a critical review. *JASIS*, 40(5):342-349, 1989.
32. MAIA, Maria de Fátima; CAREGNATO, Sônia. Estudos Bibliométricos na Comunicação Científica: Bibliotecas Digitais como Fator de Revitalização. In: ENCONTRO DE INFORMAÇÃO EM CIÊNCIAS DA COMUNICAÇÃO, 14, Porto Alegre, 2004.
33. MARCONI, Marina de Andrade; LAKATOS, Eva Maria. Metodologia do Trabalho Científico. São Paulo: Editora Atlas S.A., 1986, 198p.
34. MARCONI, Marina de Andrade; LAKATOS, Eva Maria. Técnicas de Pesquisa. São Paulo: Editora Atlas S.A., 1988, 205p.
35. MEADOWS, A. J. Communication in science. England: Butterworth & Co Ltd, 1974.
36. MEADOWS, A. J. A Comunicação científica. Brasília, DF: Briquet de Lemos, 1999.
37. MORMAN, E. T. Citation analysis and the current debate over quantitative methods in the social studies of science. *Society for the Social Studies of Science Newsletter*, 5(3), 7-13, 1981 apud CRONIN, Blaise. The citation process. London: Taylor Graham, 1984, 103p.
38. PACHECO, R. C.; KERN, V. M. Transparência e gestão do conhecimento por meio de um banco de teses e dissertações: a experiência da PPGE/UFSC. *Ciência da Informação*, Brasília, v. 30, n. 3, p. 64-72, set./dez. 2001.
39. PACKER et al. SciELO: uma metodologia para publicação eletrônica, *Ci. Inf.*, 1998, vol. 27 n° 2.
40. PROGRAMA PROSSIGA. On-line. Disponível em: <http://www.prossiga.br>. Acessado em 03/06/2005.
41. PRITCHARD, A. Statistical bibliography or bibliometrics? *J. Doc.* 25, (1969): 348-349 apud SENGUPTA, I. N. Bibliometrics, informetrics, scientometrics and librametrics: an overview. *Libri*, v. 42, n. 2, p. 75-98, 1992.
42. RAGHAVAN, Koti S. The General Theory of Classification as the Basis for Structuring of Subject Headings.

43. ROUSSEAU, Ronald. Indicadores bibliométricos e econométricos para a avaliação de instituições científicas. *Ci. Inf.*, Brasília, v. 27, n. 2, p. 149-158, maio/ago, 1998.
44. SENGUPTA, I. N. Bibliometrics, informetrics, scientometrics and librametrics: an overview. *Libri*, v. 42, n. 2, p. 75-98, 1992.
45. SHARP, John R. The SLIC index. *American Documentation*, vol. 17, 1966, p. 41-44.
46. SOERGEL, Dagobert. A Framework for Digital Library Research. *D-Lib Magazine*, v. 8, n. 12, dec/2002.
47. TARGINO, Maria das Graças; GARCIA, Joana Coeli Ribeiro. Ciência brasileira na base de dados do Institute for Scientific Information (ISI). *Ci. Inf.*, Brasília, v. 29, n. 1, p. 103-117, jan./abr. 2000.
48. VELHO, Léa Maria Strini. Estratégias para um sistema de indicadores de C&T no Brasil. *Parcerias Estratégicas*, n. 13, p. 109-121, dez/2001.
49. WEINSTOCK, M. Citation Indexes, *Encyclopedia of Library and Information Science*, volume 5, pages 16-40, New York: Marcel Drekker, Inc., 1971.
50. WERSIG, G.; NEVELLING, U. The phenomena of interest to information science. *Information Scientist*, v. 9, p. 127-140, 1975.
51. WOOSTER, H. Optimal utilization of indexing personnel. *Research Review* (U. S. Air Force, Office of Aerospace Research), 3 (4), 1964, 22-23 apud LANCASTER, F. W. *Indexação e resumos: teoria e prática*. Brasília: Briquet de Lemos, 1993.
52. ZIMAN, John. Information, Communication, Knowledge. *Nature*, v. 224, p. 318-324, 1969.
53. ZIMAN, John. *Conhecimento Público*. Belo Horizonte: Ed. Itatiaia, 1979.
54. ZIVIANI, Nivio. *Projeto de Algoritmos*. São Paulo: Pioneira Thomson Learning, 2004, 552p.

## ANEXOS

### ANEXO 1 – Lista de stop words utilizados na rotação do KWIC

A	C	DIANTE	GERALMENTE
ABAIXO	CADA	DIFICILMENTE	H
ACIMA	CATORZE	DIRETAMENTE	HA
ACOLA	CEDO	DISPONIVEL	HAYER
AGORA	CEM	DISSO	I
AI	CERTAMENTE	DISTO	INICIALMENTE
AINDA	CINCO	DO	INTEIRAMENTE
AL	CINQUENTA	DOIS	ISSO
ALEM	CLARAMENTE	DOS	ISTO
ALGO	CM	DOZE	J
ALGUEM	COM	DURANTE	JA
ALGUM	COMIGO	DUZENTOS	JAMAIS
ALGUMA	COMO	E	JR
ALGUMAS	CONCISO	ELA	JUNTO
ALGUNS	CONFORME	ELAS	K
ALI	CONOSCO	ELE	KM
ALIAS	CONSEQUENTEMENTE	ELES	L
AMBOS	CONSIGO	EM	LA
ANDAR	CONTIGO	EMBORA	LAS
ANTEONTEM	CONTRA	ENQUANTO	LO
ANTERIORMENTE	CONTUDO	ENTAO	LOGO
ANTES	CORRENTEMENTE	ENTRE	LONGE
AO	CR	ENTRETANTO	LOS
AONDE	CUJA	ESPECIALMENTE	M
AOS	CUJO	ESSA	MAIORIA
APARENTEMENTE	D	ESSAS	MAIS
APENAS	DA	ESSE	MAL
APESAR	DAQUILO	ESSES	MAS
APOS	DAS	ESTA	ME
APROPRIADO	DE	ESTAR	MEDIANTE
AQUELA	DEBAIXO	ESTAS	MELHOR
AQUELAS	DEFINITIVAMENTE	ESTE	MENOR
AQUELE	DEFRONTE	ESTES	MENOS
AQUELES	DELA	ET	MESMO
AQUI	DELE	ETC	MEU
AQUILO	DEMAIS	EU	MEUS
AS	DENTRO	EX	MIL
ASSIM	DEPOIS	EXATAMENTE	MILHAO
ATE	DESDE	EXCETO	MILHARES
ATRAVES	DESSA	EXCLUSIVAMENTE	MILHOES
ATUALMENTE	DESSAS	F	MIM
AUTOMATICAMENTE	DESSE	FACILMENTE	MINHA
B	DESSES	FAZER	MINHAS
BASICAMENTE	DESTA	FELIZMENTE	ML
BASTANTE	DESTAS	FICAR	MUITA
BEM	DESTE	FINALMENTE	MUITAS
BILHAO	DESTES	FORA	MUITO
BILHOES	DEZ	FREQUENTEMENTE	MUITOS
BONITO	DEZESEIS	G	N



NA	PERMANECER	SIM	ABOUT
NAS	PIOR	SO	ABOVE
NADA	POIS	SOB	ACCORDING
NAO	POR	SOBRE	ACROSS
NAQUILO	POREM	SP	39
NECESSARIO	PORQUANTO	SUA	ACTUALLY
NELA	PORQUE	SUAS	AD
NELE	PORTANTO	SUBITO	ADJ
NEM	POSSIVELMENTE	T	AE
NENHUM	POSTERIORMENTE	TAIS	AF
NENHUMA	POUCA	TAL	AFTER
NESSA	POUCAS	TALVEZ	AFTERWARDS
NESSAS	POUCO	TAMBEM	AG
NESSE	POUCOS	TANTA	AGAIN
NESSES	PRATICAMENTE	TANTAS	AGAINST
NESTA	PRIMEIRAMENTE	TANTO	AI
NESTAS	PRIMEIRO	TANTOS	AL
NESTE	PRINCIPALMENTE	TAO	ALL
NESTES	PRONTAMENTE	TE	ALMOST
NINGUEM	PROVAVELMENTE	TEL	ALONE
NISSE	Q	TER	ALONG
NISTO	QUAIS	TERCEIRO	ALREADY
NO	QUAL	TEU	ALSO
NONO	QUALQUER	TI	ALTHOUGH
NORMALMENTE	QUANDO	TODA	ALWAYS
NOS	QUANTA	TODAS	AM
NOSSA	QUANTAS	TODAVIA	AMONG
NOSSAS	QUANTO	TODO	AMONGST
NOSSO	QUANTOS	TODOS	AN
NOSSOS	QUARENTA	TRAS	AND
NOVAMENTE	QUASE	TRES	ANOTHER
NOVE	QUATORZE	TREZE	ANY
NOVECENTOS	QUATRO	TREZENTOS	ANYHOW
NOVENTA	QUE	TRINTA	ANYONE
NUNCA	QUEM	TU	ANYTHING
O	QUINTO	TUA	ANYWHERE
OITENTA	QUINZE	TUDO	AO
OITO	R	U	AQ
OITOCENTOS	RAPIDAMENTE	ULTIMAMENTE	AR
ONDE	REALMENTE	UM	ARE
ONTEM	S	UMA	AREN
ONZE	SE	UMAS	AREN'T
OPCIONALMENTE	SEGUNDO	UNS	AROUND
OS	SEIS	V	ARPA
OU	SEISCENTOS	VARIAS	AS
OUTRA	SEJA	VARIOS	AT
OUTRAS	SEM	VEM	AU
OUTRO	SEMPRE	VEZ	AW
OUTROS	SENAO	VEZES	AZ
P	SER	VINTE	B
PARA	SESSENTA	VOCE	BA
PARECER	SETE	W	BB
PELA	SETECENTOS	X	BD
PELAS	SETENTA	Z	BE
PELO	SEU	ZERO	BECAME
PELOS	SEUS	A	BECAUSE
PERANTE	SI	II	BECOME

BECOMES	CR	FJ	HEREBY
BECOMING	CS	FK	HEREIN
BEEN	CU	FM	HEREUPON
BEFORE	CV	FO	HERS
BEFOREHAND	CX	FOR	HERSELF
BEGIN	CY	FORMER	HIM
BEGINNING	CZ	FORMERLY	HIMSELF
BEHIND	D	FORTY	HIS
BEING	DE	FOUND	HK
BELOW	DID	FOUR	HM
BESIDE	DIDN	FR	HN
BESIDES	DIDN'T	FREE	HOW
BETWEEN	DJ	FROM	HOWEVER
BEYOND	DK	FURTHER	HR
BF	DM	FX	HT
BG	DO	G	HTM
BH	DOES	GA	HTML
BI	DOESN	GB	HTTP
BILLION	DOESN'T	GD	HU
BJ	DON	GE	HUNDRED
BM	DON'T	GET	I
BN	DOWN	GF	I'D
BO	DURING	GG	I'LL
BOTH	DZ	GH	I'M
BR	E	GI	I'VE
BS	EACH	GL	I.E.
BT	EC	GM	ID
BUT	EDU	GMT	IE
BUY	EE	GN	IF
BV	EG	GO	IL
BW	EH	GOV	IM
BY	EIGHT	GP	IN
BZ	EIGHTY	GQ	INC
C	EITHER	GR	INC.
CA	ELSE	GS	INDEED
CAN	ELSEWHERE	GT	INSTEAD
CAN'T	END	GU	INT
CANNOT	ENDING	GW	INTO
CAPTION	ENOUGH	GY	IO
CC	ER	H	IQ
CD	ES	HAD	IR
CF	ET	HAS	IS
CG	ETC	HASN	ISN
CH	EVEN	HASN'T	ISN'T
CI	EVER	HAVE	IT
CK	EVERY	HAVEN	IT'S
CL	EVERYONE	HAVEN'T	ITS
CLICK	EVERYTHING	HE	ITSELF
CM	EVERYWHERE	HE'D	J
CN	EXCEPT	HE'LL	JE
CO	F	HE'S	JM
CO.	FEW	HELP	JO
COM	FI	HENCE	JOIN
COPY	FIFTY	HER	JP
COULD	FIND	HERE	K
COULDN	FIRST	HERE'S	KE
COULDN'T	FIVE	HEREAFTER	KG

KH	MOST	ONE'S	SH
KI	MOSTLY	ONLY	SHE
KM	MP	ONTO	SHE'D
KN	MQ	OR	SHE'LL
KP	MR	ORG	SHE'S
KR	MRS	OTHER	SHOULDN
KW	MS	OTHERS	SHOULDN'T
KY	MSIE	OTHERWISE	SI
KZ	MT	OUR	SINCE
L	MU	OURS	SITE
LA	MUCH	OURSELVES	SIX
LAST	MUST	OUT	SIXTY
LATER	MV	OVER	SJ
LATTER	MW	OVERALL	SK
LB	MX	OWN	SL
LC	MY	P	SM
LEAST	MYSELF	PA	SN
LESS	MZ	PE	SO
LET	N	PER	SOME
LET'S	NA	PERHAPS	SOMEHOW
LI	NAMELY	PF	SOMEONE
LIKE	NC	PG	SOMETHING
LIKELY	NE	PH	SOMETIME
LK	NEITHER	PK	SOMETIMES
LL	NET	PL	SOMEWHERE
LR	NEVER	PM	SR
LS	NEVERTHELESS	PN	ST
LT	NEW	PR	STILL
LTD	NEXT	PT	STOP
LU	NF	PW	SU
LV	NG	PY	SUCH
LY	NI	Q	SV
M	NINE	QA	SY
MA	NINETY	R	SZ
MADE	NL	RATHER	T
MAKE	NO	RE	TAKING
MAKES	NOBODY	RECENT	TC
MANY	NONE	RECENTLY	TD
MAYBE	NONETHELESS	RESERVED	TEN
MC	NOONE	RO	TF
MD	NOR	RU	TG
ME	NOT	RW	TEST
MEANTIME	NOTHING	S	TH
MEANWHILE	NOW	SA	THAN
MG	NOWHERE	SAME	THAT
MH	NP	SB	THAT'LL
MIGHT	NR	SC	THAT'S
MIL	NU	SD	THE
MILLION	NZ	SE	THEIR
MISS	O	SEEM	THEM
MK	OF	SEEMED	THEMSELVES
ML	OFF	SEEMING	THEN
MM	OFTEN	SEEMS	THENCE
MN	OM	SEVEN	THERE
MO	ON	SEVENTY	THERE'LL
MORE	ONCE	SEVERAL	THERE'S
MOREOVER	ONE	SG	

THEREAFTER	TWO	WE'RE	WITH
THEREBY	TZ	WE'VE	WITHIN
THEREFORE	U	WELL	WITHOUT
THEREIN	UA	WERE	WON
THEREUPON	UG	WEREN	WON'T
THESE	UK	WEREN'T	WOULD
THEY	UM	WF	WOULDN
THEY'D	UNDER	WHAT	WOULDN'T
THEY'LL	UNLESS	WHAT'LL	WS
THEY'RE	UNLIKE	WHAT'S	WWW
THEY'VE	UNLIKELY	WHATEVER	X
THIRTY	UNTIL	WHEN	Y
THIS	UP	WHENCE	YE
THOSE	UPON	WHENEVER	YES
THOUGH	US	WHERE	YET
THOUSAND	USE	WHEREAFTER	YOU
THREE	USED	WHEREAS	YOU'D
THROUGH	USING	WHEREBY	YOU'LL
THROUGHOUT	UY	WHEREIN	YOU'RE
THRU	UZ	WHEREUPON	YOU'VE
THUS	V	WHEREVER	YOUR
TJ	VA	WHETHER	YOURS
TK	VC	WHICH	YOURSELF
TM	VE	WHILE	YOURSELVES
TN	VERY	WHITHER	YT
TO	VG	WHO	YU
TOGETHER	VI	WHO'D	Z
TOO	VIA	WHO'LL	ZA
TOWARD	VN	WHO'S	ZM
TOWARDS	VU	WHOEVER	ZR
TP	W	NULL	10
TR	WAS	WHOLE	Z
TRILLION	WASN	WHOM	
TT	WASN'T	WHOMEVER	
TV	WE	WHOSE	
TW	WE'D	WHY	
TWENTY	WE'LL	WILL	

## ANEXO 2 – Algoritmo Shift-And-Aproximado

```
private int shiftAndAproximado(String author1, String author2) {  
  
    String texto = rearrangeAuthorName(author1);  
    String padrao = rearrangeAuthorName(author2);  
    int maiorI = 0;  
    long[] masc = new long[256]; // tamanho do alfabeto  
    int i, j;  
    long ri, rant, rnovo;  
    int numMaxErros = calcMaxErros(texto, padrao);  
    long[] r = new long[numMaxErros + 1];  
    for (i = 0; i < 256; i++)  
        masc[i] = 0L;  
    for (i = 1; i <= padrao.length(); i++)  
        masc[padrao.charAt(i - 1)] |= 1L << (padrao.length() - i);  
    r[0] = 0L;  
    ri = 1L << (padrao.length() - 1L);  
    for (j = 1; j <= numMaxErros; j++)  
        r[j] = (1L << (padrao.length() - j)) | r[j - 1];  
    for (i = 0; i < texto.length(); i++) {  
        rant = r[0];  
        rnovo = ((rant >> 1L) | ri) & masc[texto.charAt(i)];  
        r[0] = rnovo;  
        for (j = 1; j <= numMaxErros; j++) {  
            rnovo = ((r[j] >> 1L) & masc[texto.charAt(i)]) | rant | ((rant | rnovo) >> 1L);  
            rant = r[j];  
            r[j] = rnovo | ri;  
        }  
        if ((rnovo & 1L) != 0) {  
            maiorI = i + 1;  
        }  
    }  
    return maiorI;  
}
```

### ANEXO 3 – Resultado do KWIC nos títulos dos documentos analisados

<p>MACOES DO RELATORIO ANUAL DAS COMPANHIAS DE CAPITAL MACAO E O CONHECIMENTO NA BIOLOGIA DO CONHECER: UMA APLICACAO DA S DE INFORMACAO EM SAUDE DO BRASIL E DA FRANCA: UMA O USO DA INTERNET COMO FONTE DE PESQUISA</p> <p>TADO: ESTUDO EXPLORATORIO SOBRE POLITICA PUBLICA DE O USO DA INFORMACAO PARA A ALTA DIVULGACAO DA LITERATURA DESTINADA A CRIANCAS E</p> <p>O VALOR A REALIDADE DO INFORMACAO, APRENDIZAGEM, CONHECIMENTO; A TRIPLICE O USO DA INFORMACAO PARA A COMO FONTE DE PESQUISA ACADEMICA POR PROFESSORES E MONITORACAO MONITORACAO</p> <p>PRESAS [MANUSCRITO]: A UTILIZACAO DE INFORMACOES DO ESCOLA E A FORMACAO DO SUJEITO-LEITOR [MANUSCRITO]: FATORES INTERFERENTES NO PROCESSO DE O DA INFORMACAO E DO CONHECIMENTO NAS ORGANIZACOES: UMA</p> <p>IAS DE CAPITAL ABERTO: UMA AVALIACAO SOB A OTICA DE COMPUTACAO E DA INFORMACAO, UTILIZANDO-SE O METODO O VALOR AGREGADO DAS INFORMACOES DO RELATORIO</p> <p>IDADE/EMPRESA NA POS-GRADUACAO: CONTRIBUICAO PARA A INFORMACAO, IMPORTANCIA DA INFORMACAO PARA OS PROFISSIONAIS DA UDO DO USO DE PERIODICOS CIENTIFICOS BRASILEIROS DA CAO EM UMA BIBLIOTECA UNIVERSITARIA POR USUARIOS DA RTIR DOS REFERENCIAIS DA CIENCIA DA INFORMACAO E DA S: ESTUDO DE DEFINICOES PRESENTES NA LITERATURA DAS UMA ANALISE DA TEORIA DOS FANCIA DA INFORMACAO PARA OS PROFISSIONAIS DA AREA- FATORES INTERFERENTES NO PROCESSO DE ANALISE DE CONSTRUCAO DA CIDADANIA: UM ESTUDO DE CASO SOBRE OS MREBAS: O EMPREGO DAS TECNOLOGIAS DE INFORMACAO NA</p>	<p>ABERTO: UMA AVALIACAO SOB A OTICA DE ANALISTAS DE FUNDOS DE PENSAO. ABORDAGEM COGNITIVA PARA OS ESTUDOS SOBRE INTELIGENCIA EMPRESARIAL ABORDAGEM SENSE-MAKING NO ESTUDO DO COMPORTAMENTO INFORMACIONAL DE PESQ ABORDAGEM A PARTIR DOS REFERENCIAIS DA CIENCIA DA INFORMACAO E DA AREA ACADEMICA POR PROFESSORES E ALUNOS DO PROGRAMA DE POS-GRADUACAO EM CIEN ACESSO A INFORMACAO GOVERNAMENTAL: ESTUDO EXPLORATORIO DO PAPEL DO SIAF ACESSO A INFORMACAO GOVERNAMENTAL ADMINISTRACAO DAS INSTITUICOES FINANCEIRAS BANCARIAS NO BRASIL. ADOLESCENTES: A PRATICA DE EDITORAS MINEIRAS AGENCIAS DE NOTICIAS NA INTERNET COMO SERVICOS DE INFORMACAO PARA NEGOC AGREGADO DAS INFORMACOES DO RELATORIO ANUAL DAS COMPANHIAS DE CAPITAL A AGRICULTOR FAMILIAR DO MUNICIPIO DE GOVERNADOR VALADARES: SUAS NECESSID ALIANCA NO CONTEXTO DA EDUCACAO A DISTANCIA ALTA ADMINISTRACAO DAS INSTITUICOES FINANCEIRAS BANCARIAS NO BRASIL. ALUNOS DO PROGRAMA DE POS-GRADUACAO EM CIENCIA DA INFORMACAO DA UFNG AMBIENTAL E O PROCESSO DECISORIO EM PEQUENAS EMPRESAS [MANUSCRITO]: A U AMBIENTAL: UM ESTUDO DE CASO EM EMPRESAS. AMBIENTE. ANALISANDO A PRODUCAO DE LEITURA NO ENSINO MEDIO DO CENTRO FEDERAL DE E ANALISE DE ASSUNTO: ESTUDO DE CASO DE INDEXADORES ANALISE DE CASOS RELATADOS EM ORGANIZACOES PUBLICAS E PRIVADAS ANALISE DA TEORIA DOS ARQUIVOS ANALISTAS DE FUNDOS DE PENSAO. ANALITICO-SINTETICO ANUAL DAS COMPANHIAS DE CAPITAL ABERTO: UMA AVALIACAO SOB A OTICA DE AN APLICACAO DA ABORDAGEM SENSE-MAKING NO ESTUDO DO COMPORTAMENTO INFORMAC APLICACOES DE SOFTWARES DE GESTAO DO CONHECIMENTO: TIPOLOGIA E USOS. APRENDIZAGEM, A GESTAO DO CONHECIMENTO E A INOVACAO NA INDUSTRIA MINEIR APRENDIZAGEM, CONHECIMENTO; A TRIPLICE ALIANCA NO CONTEXTO DA EDUCACAO AREA-ARTISTICO-CULTURAL: UM ESTUDO EXPLORATORIO NA CIDADE DE BELO HORIZ AREA DE FISICA POR DOCENTES DOS PROGRAMAS DE POS-GRADUACAO DO NORDESTE. AREA DE ODONTOLOGIA AREA DE SAUDE. AREAS DAS CIENCIAS DA COMPUTACAO E DA INFORMACAO, UTILIZANDO-SE O METOD ARQUIVOS ARTISTICO-CULTURAL: UM ESTUDO EXPLORATORIO NA CIDADE DE BELO HORIZONTE ASSUNTO: ESTUDO DE CASO DE INDEXADORES ATORES SINDICAIS DA REDE MUNICIPAL DE ENSINO DE BELO HORIZONTE. AUTO COMERCIAL DIAMANTINA S A</p>
---	--

## ANEXO 4 – Formulário do Dspace de adição de um documento – 1ª etapa

Descrever **Descrever** Descrever Carregar Verificar Licença Concluir

**Depósito: Descreva o Seu Registo**

Por favor, introduza a informação necessária acerca do seu Depósito. Na maioria dos browsers pode usar a tecla TAB para navegar no formulário. ([Ajuda...](#))

Introduza em baixo os nomes dos autores.

Último Nome  
ex: *Silva*

Primeiro(s) nome(s)  
ex: *Manuel*

**Autores**   Adicionar Mais

Introduza o título principal do registo.

**Título**

Introduza o nome do(a) Orientador(a).

**Orientador**

Forneça a data da publicação ou distribuição pública. Pode deixar o mês e/ou o dia em branco se não se aplicar.

**Data** Mês:  (em branco)  Dia:  Ano:

Introduza a citação para a publicação anterior deste item.

**Citação**

Selecione o tipo do conteúdo que está a depositar.

**Tipo**

Selecione a língua do conteúdo principal do item. Se a língua não aparecer na lista em baixo seleccione "Outra"

**Língua**

< Anterior Próximo > Cancelar/Gravar

## ANEXO 5 – Formulário do Dspace de adição de um documento – 2ª etapa

Descrever Descrever **Descrever** Carregar Verificar Licença Concluir

**Depósito: Descreva o Registo**

Introduza informação adicional em baixo. ([Ajuda...](#))

Introduza as palavras chave ou frases do assunto.

**Palavras Chave**

Introduza as Referências Bibliográficas do documento.

**Referências**

Introduza em baixo o resumo (abstract) no idioma principal do registo.

**Resumo**

Introduza em baixo o resumo (abstract) noutra idioma.

**Resumo**

Introduza os nomes de eventuais patrocinadores na caixa em baixo.

**Patrocinadores**

Descrição adicional e comentários.

**Descrição**



## ANEXO 6 – Formulário do Dspace de adição de um documento – 3ª etapa


Descrever Descrever Descrever Carregar Verificar Licença Concluir

**Depósito: Carregar Ficheiro**

Introduza o nome do ficheiro no seu disco local correspondente ao seu registo. Se seleccionar "Procurar...", será aberta uma nova janela com a qual poderá localizar e seleccionar o ficheiro do seu disco local. ([Ajuda...](#))

**Utilizadores de Netscape** : Por defeito, a janela aberta com "Procurar..." mostra apenas ficheiros do tipo HTML. Se o seu ficheiro não é do tipo HTML seleccione na janela a opção para mostrar ficheiros de outros tipos. [Instruções para Utilizadores de Netscape](#)

Note que o sistema DSpace é capaz de preservar melhor determinados tipos de ficheiros em relação a outros. [Informação acerca de tipos de ficheiros](#) e níveis de suporte.

Ficheiro:  

< Anterior Próximo > Cancelar/Gravar

## ANEXO 7 – Formulário de cadastro das variáveis dependentes de um documento

Edit Reference of an Item			
<p><b>PLEASE NOTE: These changes are not validated in any way. You are responsible for entering the data in the correct format. If you are not sure what the format is, please do NOT make changes.</b></p>			
<p><b>Item internal ID:</b> 101</p>			
<p><b>Handle:</b> 123456789/60</p>			
<p><b>Last modified:</b> 8-Nov-2004 16:35:47</p>			
<p><b>Item page:</b> <a href="http://localhost/dspace/handle/123456789/60">http://localhost/dspace/handle/123456789/60</a></p>			
Reference			
<p>1) ABRAMO, Claudio Weber. Acesso a informacao: um longo caminho a ser trilhado. In: KONDO, Seiichi (et al.). Transparencia e responsabilizacao no setor publico: fazendo acontecer. Brasilia: MP/SEGES, 2002. Disponível em Acessado em: 13/02/2003</p>			
Country:	Type:	Language:	Date:
Brazil	Livro	Portuguese	2002
Author:	Title:		
ABRAMO, Claudio We	Acesso a informacao: um longo caminho a ser trilhado		
<p>2) ABRANCHES, Sergio. O Estado. In: JAGUARIBE, Helio et al. Sociedade, Estado e partidos: na atualidade brasileira. Rio de Janeiro: Paz e Terra, 1992.</p>			
Country:	Type:	Language:	Date:
Brazil	Livro	Portuguese	1992
Author:	Title:		
ABRANCHES, Sergio	O Estado		
<p>3) ABRUCIO, Fernando Luiz. O impacto do modelo gerencial na administracao publica: um breve estudo sobre a experiencia internacional recente. Brasilia: ENAP, 1997. 52 p. (Cadernos ENAP, n.10)</p>			
Country:	Type:	Language:	Date:
Brazil	Livro	Portuguese	1997
Author:	Title:		
ABRUCIO, Fernando L	O impacto do modelo gerencial na administracao publica: um breve estudo s		
<p>4) ABRUCIO, Fernando Luiz. Os avancos e os dilemas do modelo pos-burocratico: a reforma da administracao publica a luz da experiencia internacional recente. In: BRESSER PEREIRA, Luiz Carlos; SPINK, Peter (Orgs.). Reforma do Estado e administracao publica gerencial. 2 ed. Rio de Janeiro: Fundacao Getulio Vargas, 1998.</p>			
Country:	Type:	Language:	Date:
Brazil	Livro	Portuguese	1998
Author:	Title:		
ABRUCIO, Fernando L	Os avancos e os dilemas do modelo pos-burocratico: a reforma da administr		
<p>5) ALLEN, Bryce. Information tasks: toward a user centered approach to information systems. San Diego: Academic Press, 1996.</p>			
Country:	Type:	Language:	Date:
United States	Livro	English	1996
Author:	Title:		
ALLEN, Bryce	Information tasks: toward a user centered approach to information systems		
<p>6) ALMEIDA, Jorge. Midia, opiniao publica ativa e esfera publica democratica. Anais IV Congresso Latinoamericano de ciencias de la comunicacion ALAIC. 12 a 16 de setembro de 1998. Disponível: Acessado em: 09/08/2002.</p>			
Country:	Type:	Language:	Date:
Brazil	Trabalho apresentado em Congresso	Portuguese	1998

ANEXO 8 – Formulário de criação dos arquivos de autoridade dos nomes dos autores

Reference	
1) Reference: 240928	Author name: <input checked="" type="radio"/> 3278 - ALBAGLI, S
<a href="#">Manual</a>	
<input type="radio"/> 42 - ALBAGLI, S <input type="radio"/> 3280 - ALBAGLI, Sarita <input type="radio"/> 618 - BLILI, S <input type="radio"/> 6814 - LI, F	
2) Reference: 237188	Author name: <input checked="" type="radio"/> 42 - ALBAGLI, S
<a href="#">Manual</a>	
<input type="radio"/> 3278 - ALBAGLI, S <input type="radio"/> 3280 - ALBAGLI, Sarita <input type="radio"/> 618 - BLILI, S <input type="radio"/> 6814 - LI, F	
3) Reference: 243758	Author name: <input checked="" type="radio"/> 3280 - ALBAGLI, Sarita
<a href="#">Manual</a>	
<input type="radio"/> 3278 - ALBAGLI, S <input type="radio"/> 3282 - ALBAGLI, Sarita <input type="radio"/> 618 - BLILI, S <input type="radio"/> 6814 - LI, F	

ANEXO 9 – Formulários de criação dos arquivos de autoridade dos assuntos dos títulos dos documentos e das referências

[DSpace at My University](#) >  
[Administer](#) >

### Edit Clusters of Subjects

**PLEASE NOTE: These changes are not validated in any way. You are responsible for entering the data in the correct format. If you are not sure what the format is, please do NOT make changes.**

Choose the initial letter:

[A](#) | [B](#) | [C](#) | [D](#) | [E](#) | [F](#) | [G](#) | [H](#) | [I](#) | [J](#) | [K](#) | [L](#) | [M](#) | [N](#) | [O](#) | [P](#) | [Q](#) | [R](#) | [S](#) | [T](#) | [U](#) | [V](#) | [W](#) | [X](#) | [Y](#) | [Z](#)

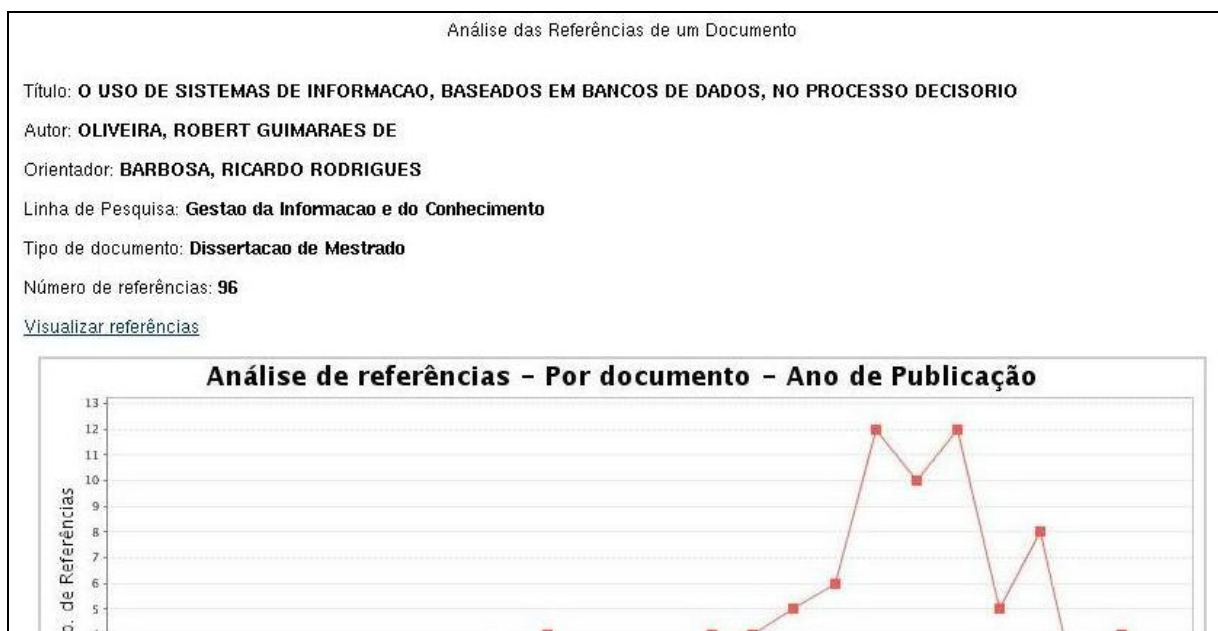
**Reference**

### Edit Clusters of Subjects

**PLEASE NOTE: These changes are not validated in any way. You are responsible for entering the data in the correct format. If you are not sure what the format is, please do NOT make changes.**

Reference		
1) Item: 99	Subject name: <input checked="" type="radio"/> 2 - ABORDAGEM	Another subject: <input type="text"/>
<a href="#">Manual</a>		
<input type="radio"/> 3 - ABORDAGEM <input type="radio"/> 574 - ABORDAGENS		
2) Item: 90	Subject name: <input type="radio"/> 3 - ABORDAGEM	Another subject: <input type="text"/>
<a href="#">Manual</a>		
<input checked="" type="radio"/> 2 - ABORDAGEM <input type="radio"/> 574 - ABORDAGENS		
3) Item: 55	Subject name: <input type="radio"/> 4 - ABORDAGEM	Another subject: <input type="text"/>
<a href="#">Manual</a>		
<input checked="" type="radio"/> 2 - ABORDAGEM <input type="radio"/> 574 - ABORDAGENS		

## ANEXO 10 – Tela de análise bibliométrica para um único documento – 1ª Etapa - Gráficos



## ANEXO 11 – Tela de análise bibliométrica para um único documento – 2ª Etapa - Referências

**Visualizar referências de um documento**

**Item ID:** 26  
**Handle:** 123456789/25  
**Última alteração:** 8-Mar-2005 19:15:05  
**Acesse o documento:** <http://localhost/dspace/handle/123456789/25>

Ordernar por: [País](#) | [Tipo de Publicação](#) | [Idioma](#) | [Ano](#) | [Autor](#)

Referências			
1) ABREU, Aline F. Gestao tecnologica e planejamento estrategico.Convenio. IDAQ. Sao Paulo, 1996.			
<b>País:</b> <b>Brazil</b>	<b>Tipo de publicação:</b> <b>Livro</b>	<b>Idioma:</b> <b>Portuguese</b>	<b>Ano:</b> <b>1996</b>
<b>Autor:</b> <b>ABREU, Aline F</b>	<b>Título:</b> <b>Gestao tecnologica e planejamento estrategico</b>		
2) ALBERTIN, Alberto Luiz. Aumentando as chances de sucesso no desenvolvimento e implementacao de sistemas de informacao. Revista de Administracao de Empresas, Sao Paulo, v. 36, n.3, 1996.			
<b>País:</b> <b>Brazil</b>	<b>Tipo de publicação:</b> <b>Artigo de Periódico</b>	<b>Idioma:</b> <b>Portuguese</b>	<b>Ano:</b> <b>1996</b>
<b>Autor:</b> <b>ALBERTIN, Alberto Luiz</b>	<b>Título:</b> <b>Aumentando as chances de sucesso no desenvolvimento e implementacao de sistemas</b>		