**INSTITUTO DE PESQUISAS ENERGÉTICAS E
NUCLEARES**
Autarquia Associada à Universidade de São Paulo

**Elaboração de indicadores de produção científica com base na
análise cientométrica das dissertações e teses do IPEN**

Mery Piedad Zamudio Igami

Tese apresentada como parte dos
requisitos para obtenção do Grau de
Doutor em Ciências da Área de
Tecnologia Nuclear – Materiais

Orientador:

Prof. Dr. José Carlos Bressiani

São Paulo
2011

Dedico
ao Ruy, à Thais e ao Daniel, pela maravilhosa família que formamos.
Vocês são um presente de Deus!

*Antes do compromisso há hesitação,
a oportunidade de recuar, a ineficácia permanente.*

*Porém no momento em que nos comprometemos de fato,
a providência age também. Ocorre toda espécie de coisas para nos
ajudar, coisas que de outro modo nunca ocorreriam.*

*Toda uma cadeia de eventos surge da decisão fazendo vir em
nosso favor todo tipo de encontros, de incidentes e de apoio material
imprevistos que ninguém poderia sonhar que viria em seu caminho.*

*Comece tudo o que pode fazer ou que sonha que pode fazer,
há gênio, poder e mágica na ousadia.*

(Goethe, 1887)

AGRADECIMENTOS

A elaboração de um trabalho como este é uma experiência enriquecedora em todos os aspectos. É um desafio pessoal, profissional, intelectual; porém, há um aspecto marcante durante esta construção: a oportunidade de interagir com pessoas que de alguma forma, em menor ou maior grau, contribuíram para a concretização da tese. É com muita satisfação que registro meu reconhecimento a estas pessoas.

Inicio pelo meu profundo reconhecimento e agradecimento ao meu Orientador, Dr. José Carlos Bressiani, por me propor o “desafio” inicial de realizar o doutorado no Programa do IPEN. Obrigada pela confiança, pelo apoio, pelo acompanhamento e pelas críticas construtivas.

Aos Professores, Dra. Nair Yumiko Kobashi, da ECA/USP, e Dr. Raimundo Nonato dos Santos, da UFR, pelas conversas, pelos incentivos e ensinamentos iniciais, quando o trabalho ainda se constituía em uma “ideia possível e promissora”.

Ao Dr. José Roberto Rogero, ex-diretor no IPEN, com quem tive a oportunidade de trabalhar, e a quem devo o apoio e o incentivo inicial para completar este ciclo acadêmico. A palavra amiga, a atenção e a disponibilidade para a troca de ideias e sugestões foram fundamentais para que eu prosseguisse na elaboração do trabalho. Meu respeito e agradecimento.

Agradeço e também registro meu reconhecimento à excelente logística e qualidade de serviços disponibilizados para os alunos do Programa de Pós-graduação do IPEN, esta qualidade é fruto de uma gestão eficiente e do desempenho de uma equipe capacitada, o que proporciona tranquilidade ao aluno para realizar seu trabalho.

Ao contrário de que se pensa a elaboração de uma tese não é um mérito individual, e sim um trabalho que requer colaboração, auxílio, convergência de ideias e atitudes positivas para realizar os processos de construção desse trabalho.

No entanto, nominar cada uma destas pessoas alongaria por demais estas páginas, além de correr o risco de cometer injustiças involuntárias. Confio na intuição de que cada uma delas sabe da minha enorme gratidão e saberá se reconhecer nestas simples palavras de agradecimento.

Assim, agradeço aos autores das teses que atenciosamente e em tempo recorde se dispuseram a responder aos questionários, validando uma etapa importante do trabalho.

Aos pesquisadores seniores que gentilmente concordaram em trocar ideias e que me permitiram registrar suas informações.

Aos colegas de disciplinas do Programa, aos companheiros de trabalho da biblioteca e de fora dela, a todos os amigos “interlocutores” que se dispuseram a discutir, sugerir, fornecer informações, extrair, testar e coletar dados, e sobretudo àqueles que souberam “ouvir pacientemente” minhas reflexões e incertezas acolhendo-me com amizade.

Aos integrantes da banca, tanto titulares como suplentes, que tão pronta e gentilmente responderam ao meu convite.

Agradeço a paciência e amizade de todos os profissionais com cujos serviços tive a sorte de contar e que com muita dedicação e competência me auxiliaram durante o tempo de processamento e tratamento dos dados.

ELABORAÇÃO DE INDICADORES DE PRODUÇÃO CIENTÍFICA COM BASE NA ANÁLISE CIENTOMÉTRICA DAS DISSERTAÇÕES E TESES DO IPEN

Mery Piedad Zamudio Igami

RESUMO

A importância da elaboração de indicadores está amplamente difundida na sociedade atual. Observa-se seu uso rotineiro nos mais diversos segmentos da sociedade. Na área científica essa prática também já está consolidada, corrobora esta afirmação o número de trabalhos registrados nas principais bases de dados; prevalecem, no entanto, os estudos quantitativos, os quais fazem uso de dados obtidos a partir de bases de dados internacionais. Diante dessa constatação o objetivo deste trabalho foi obter indicadores científicos a partir da análise das dissertações e teses produzidas por um programa de pós-graduação no período de 1977 até 2009. Foram obtidos três tipos de indicadores bidimensionais, numéricos temáticos e de produtividade, para tanto foram utilizadas técnicas de bibliometria avançada. As dissertações e teses foram categorizadas tematicamente utilizando a *Subject Categories and Scope Descriptions* e o Tesauro do *International Nuclear Information System* (INIS). Por meio da técnica de análise de *clusters* e de trajetória, as dissertações e teses foram agrupadas em cinco grupos temáticos demonstrando o comportamento passado e a tendência de crescimento de cada grupo. Para a extração dos dados sobre a produtividade das teses, em nível de artigos, foi utilizado o Currículo Lattes e a base local do IPEN. Especificamente para se identificarem os níveis de correlação entre os artigos e as teses, foi utilizada a técnica de coocorrência de palavras (*co-word analysis*) e no caso foram utilizadas as palavras-chave de ambos os itens de literatura. Os autores das teses foram consultados, por meio de um questionário, para validação dos dados. Obtiveram-se indicativos de que 55,9% dos artigos produzidos em um período de cinco anos antes e cinco anos depois da defesa da tese têm forte correlação com esta. Foi observado também que os artigos correlacionados têm sido publicados, em média, 1,63 (ano) antes da apresentação da tese. Foi demonstrado, que as bases de dados, quando bem estruturadas, são instrumentos multifuncionais com grande potencial para o uso em estudos dessa natureza. Observou-se que a qualidade dos indicadores está fortemente associada à qualidade dos dados obtidos das bases de dados.

Palavras-chave: cientometria; bibliometria; dissertações; teses; indicadores; produção científica; coocorrência; energia nuclear, bases de dados.

CONSTRUCTION OF SCIENTIFIC PRODUCTION INDICATORS BASED ON SCIENTOMETRICS ANALYSIS OF IPEN DISSERTATIONS AND THESES

Mery Piedad Zamudio Igami

ABSTRACT

Construction of Indicators plays an important role in the contemporaneous society. It could be observed their ordinary use by all the activities segments. In scientific area it is not different; such practice has already been consolidated; this statement could be corroborated by the innumerable papers published about this matter in the main databases; however, there is a prevalence of quantitative studies, which obtain data from the international databases, analyzing journal articles. Concerning this finding, the main objective of this study was to elaborate scientific indicators from a local institutional data base, using as a *corpus* of the study the dissertations and theses, produced by a graduate program in the period of 1977 to 2009. Three types of two-dimensional indicators were obtained by using bibliometric techniques: numerical, thematic and productivity. For dissertations and theses thematic characterization it was used the *Subject categories and scope descriptions* and the *International Nuclear Information System Thesaurus* (INIS). Through the statistical technique of clustering analysis, it was possible to classify them in five main groups, showing former performance and growth future trends of each group; for data extraction about theses productivity, regarding articles published; the Curriculo Lattes, and the local institutional database were used. Co-word analysis technique was used to establish a more precise correlation, between articles and theses, and for this purpose it was used the keywords from a controlled vocabulary. In order to validate the results, it was performed a survey, with the theses authors. Results obtained indicated that 55.9% of the articles published, between 5 years, before and after the thesis presentation, are strongly correlated to it. It was also observed that, correlated articles have been published, in an average of, 1, 63 year before the thesis presentation. Concerning databases, it was shown that they are valuable tools and could be used with multifunctional purposes, with a great potential to be used in bibliometric studies. It was also observed that the indicators reliability is strongly associated to the quality of data obtained from the databases.

Keywords: scientometrics; bibliometric; dissertations; theses; indicators; scientific production; co-word analysis, nuclear energy, databases.

SUMÁRIO

1 INTRODUÇÃO	12
1.1 Objetivos	15
1.1.1 Geral	15
1.1.2 Específicos	15
1.2 Justificativa	15
1.3 Apresentação e organização deste trabalho	21
2 O INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES, IPEN-CNEN/SP	23
2.1 A área nuclear no Brasil	24
2.1.1 A institucionalização da área nuclear e o Programa de Pós-graduação	35
3 QUADRO TEÓRICO	37
3.1 A produção de indicadores em ciência	46
3.1.1 Histórico dos indicadores	47
3.1.2 Os mapas da ciência	57
3.1.3 A técnica de coocorrência de palavras (<i>co-word analysis</i>)	58
3.2. A bibliometria e a cientometria e a informetria	59
3.3 A literatura cinzenta (LC) e as Dissertações e Teses	66
3.4 O uso das teses como <i>corpus</i> de pesquisa	71
3.5 As bases de dados (BDs)	72
3.5.1 Os repositórios de informação	80
3.5.2 Os sistemas de recuperação da informação	85
3.5.3 A base de dados da Produção Científica do IPEN (PTC)	89
3.5.4 Os procedimentos utilizados para a indexação dos documentos	89
4 METODOLOGIA	92
4.1 Procedimentos adotados para construção da base de dados das dissertações e teses	95
4.2 A produtividade das teses com relação à publicação de artigos	96

4.3	Procedimentos adotados para a construção das tabelas matrizes dos artigos ...	98
4.4	A técnica de coocorrência.....	101
4.5	Tempo para publicação e a visibilidade dos artigos	105
5	RESULTADOS	107
5.1	Distribuição temporal e temática das dissertações e teses	107
5.2	Trajetória e tendências temáticas.....	113
5.3	Resultados da verificação dos artigos.....	119
5.3.1	Resultados da consulta aos autores das teses.....	120
5.3.2	Cálculo do intervalo de confiança.....	123
5.3.3	Aplicação do método de coocorrência no universo total de artigos.....	126
5.3.4	Resultado da verificação do tempo para publicação dos artigos.....	135
6	DISCUSSÃO	136
6.1	A análise das séries temporais	136
6.2	A correlação dos temas com a história do Programa de Pós-graduação	140
6.3	A produtividade das teses com relação aos artigos publicados	144
6.4	O Currículo Lattes	149
6.5	As fontes de informação.....	150
6.6	O potencial das bases de dados para estudos bibliométricos	151
7	CONCLUSÃO	155
7.1	Limitações do estudo e sugestões para futuros trabalhos.....	157
	APÊNDICE A – Exemplo do questionário enviado para os autores das teses	159
	APÊNDICE B – Tabela matriz construída a partir das informações extraídas do Currículo Lattes e da PTC.....	160
	REFERÊNCIAS BIBLIOGRÁFICAS	161

LISTA DE FIGURAS

Figura 1 – O <i>continuum</i> do conhecimento.	19
Figura 2 – Estrutura adotada para a elaboração deste trabalho	22
Figura 3 – Número de doutores titulados no Brasil 1987-2008	43
Figura 4 – Número de trabalhos e seu percentual em relação aos trabalhos científicos no mundo	44
Figura 5 – Crescimento da titulação de doutores e número de artigos publicados	45
Figura 6 – Manuais metodológicos da Organization for Economic Co-operation and Development (OCDE)	50
Figura 7 – Tipos de nível de indicadores	55
Figura 8 – Fluxo simplificado do funcionamento e da coleta da Produção Técnico Científica digital no âmbito do Instituto.....	91
Figura 9 – Representação gráfica dos procedimentos envolvidos em um estudo utilizando os métodos bibliométricos para análise da informação.....	93
Figura 10 – Etapas a serem seguidas para se elaborar um estudo bibliométrico	94
Figura 11 – Exemplo de correlação entre o número de palavras-chave (PCs) e o número de coocorrências.....	103
Figura 12 – Exemplo de coocorrência de palavras-chave entre uma tese e um artigo	104
Figura 13 – Distribuição temporal das dissertações e teses defendidas no Programa de Pós-Graduação do IPEN no período de 1977 a 2009.....	107
Figura 14 – Distribuição numérica das dissertações e teses, com as categorias temáticas nucleares agrupadas, período 1977-2009	109
Figura 15 – Distribuição percentual das dissertações e teses, mantendo-se as categorias temáticas nucleares agrupadas, período 1977-2009.....	110
Figura 16 – Distribuição temática, percentual das 10 categorias com maior número de dissertações e teses, sem agrupamento temático, período 1977-2009.....	111
Figura 17 – Distribuição temática das dissertações e teses em 17 categorias no período inicial do programa de PG, período 1977-1984	112
Figura 18 – Distribuição temática das dissertações e teses em 15 categorias no período 1985-1989.....	113

Figura 19 – Distribuição temática das dissertações e teses em 21 áreas, no último período estudado 2005-2009	113
Figura 20 – Evolução e tendências dos 5 grupos temáticos formados pelas dissertações e teses, após a aplicação da técnica de análise de clusters, distribuídas pelos quinquênios estudados 1977-2009.....	116
Figura 21 – Visualização gráfica do nível de produção das séries temáticas distribuídas em quinquênios.....	116
Figura 22 – Resumo, em porcentagem, dos resultados das respostas da opinião dos autores com relação à correlação dos artigos.....	120
Figura 23 – Resultados comparativos entre o teste automático de coocorrência de palavras-chave e os questionários respondidos pelo autor	121
Figura 24 – Resumo dos resultados comparativos entre o teste automático e a opinião do autor	122
Figura 25 – Resultados obtidos, em percentual, após a aplicação do teste automático no universo total de artigos (2211) produzidos pelas teses no período analisado 1982-2009	126
Figura 26 – Distribuição temporal dos artigos, pelo ano de publicação da tese e pelo ano de publicação do artigo, sem a aplicação da técnica de correlação	128
Figura 27 – Distribuição do nº total de artigos, correlacionados, publicados em periódicos internacionais, e indexados na WoS no período estudado, 1982-2009 ...	134
Figura 28 – Distribuição comparativa dos resultados evidenciando o nº de teses, a diferença entre o nº total de artigos e o nº de artigos correlacionados, e a linha de tendência.....	136
Figura 29 – Média de tempo, em anos, para publicação dos artigos correlacionados (1237), em relação à data de apresentação da tese	135

LISTA DE TABELAS

Tabela 1 – Evolução do Fator de Impacto dos 10 títulos de periódicos que mais produziram artigos no período de 2001 a 2009	106
Tabela 2 – Distribuição das Dissertações e Teses produzidas pelo Programa de Pós-Graduação do IPEN no período de 1977 a 2009 por categorias temáticas	108
Tabela 3 – Resultado do agrupamento temático das dissertações e teses, após a análise de trajetória e análise de clusters	115
Tabela 4 – Resultados da análise hierárquica de Ward, das categorias temáticas das dissertações e teses, período 1977-2009 com classificação de níveis	118
Tabela 5 – Distribuição das teses e dos artigos publicados e declarados pelos autores das teses no Currículo Lattes, distribuídos por ano de defesa da tese, sem a aplicação da técnica de coocorrência.	129
Tabela 6 – Distribuição temporal dos artigos declarados pelos autores das teses coletados no Currículo Lattes.....	130
Tabela 7 – Distribuição das teses, do nº total de artigos declarados pelos autores das teses e o nº de artigos correlacionados publicados em periódicos internacionais e indexados na WoS.	132
Tabela 8 – Nº de teses, média do nº total de artigos, e média do número de artigos correlacionados no período intermediário analisado, 1985-2004.....	132
Tabela 9 – Média total de artigos produzidos e declarados pelos autores das teses e média de artigos correlacionados com a tese após a aplicação da técnica de coocorrência, distribuídos pelo ano de defesa da tese, período 1982-2009	133

1 INTRODUÇÃO

As atividades de ciência e tecnologia (C&T) adquiriram significativa importância na sociedade contemporânea devido em parte à grande influência que exercem no desenvolvimento econômico, político e cultural dos países. Não podem ser estudadas fora do contexto social em que estão inseridas; convivem em um mesmo espaço, em outras palavras, atuam em benefício mútuo. Ao longo da história, a função social da ciência sob diversas perspectivas tem sido exaustivamente abordada por vários autores (BUSH (1945); BERNAL (1973); MERTON (1973). A *United Nations, Educational, Scientific and Cultural Organization*(UNESCO), em sua *declaração sobre a ciência e o uso do conhecimento científico* apresentada na Conferência Geral realizada em Paris, sustenta:

... a função inerente da atividade científica é realizar pesquisas abrangentes e exaustivas da natureza e da sociedade que conduzam a um novo conhecimento. Esse novo conhecimento deve proporcionar um enriquecimento educacional, cultural e intelectual bem como propiciar benefícios sociais e avanços tecnológicos. Promover pesquisas fundamentais ou tecnológicas é essencial para alcançar um desenvolvimento e progresso endógeno... hoje mais do que nunca não pode haver desenvolvimento sem a ciência e suas aplicações (UNITED NATIONS , EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION,1999) (tradução livre da autora).

Seja na pesquisa básica, seja no desenvolvimento de bens e produtos para o bem-estar da sociedade, observam-se, atualmente, no cotidiano, os benefícios dos resultados da pesquisa científica de forma mais veloz que em épocas anteriores, devido ao expressivo desenvolvimento das tecnologias de comunicação. Assim, sendo a atividade científica elemento básico no desenvolvimento da sociedade, torna-se necessário conhecer os resultados dos recursos que a sociedade destina a essa atividade.

A evolução da ciência se materializa na produção científica por meio dos suportes documentais, tendo em vista que a pesquisa, uma vez realizada, precisa

ser comunicada para ser socializada; portanto, a comunicação e a difusão da ciência, por meio de documentos escritos, são os elementos essenciais da prática científica. A pesquisa é estimulada e se sustenta pelo fluxo de novas informações. Quando um ciclo de informação se completa, gera-se uma nova informação, numa interação contínua e infinita, formando-se um ciclo constante de criação e descobrimento (SPINAK, 1998).

A ciência, então, alimenta-se da própria ciência; diz-se ainda que a ciência é uma atividade cumulativa, sequencial e coletiva. A atividade científica deve ser vista, interpretada e medida no contexto político, social, econômico e histórico em que está inserida; isso significa que a ciência não pode ser medida em uma escala absoluta, e sim em relação às expectativas da sociedade em que está inserida (MACIAS-CHAPULA, 1998).

Com base na premissa que classifica a pesquisa científica como uma atividade social, sua avaliação de desempenho vai se tornando cada dia mais importante. Faz-se necessário acompanhar esse desenvolvimento por meio de instrumentos de mensuração e avaliação da atividade científica e seu impacto na sociedade, com o principal objetivo de adequar convenientemente a distribuição de recursos orçamentários destinados à pesquisa e ao desenvolvimento. Conhecer o perfil, medir e avaliar a produção científica são, desse modo, aspectos importantes da gestão e do planejamento científico de qualquer Instituição dedicada à pesquisa.

Pela mensuração são produzidos indicadores científicos, baseados em análises estatísticas de produção científica, obtidos pelas técnicas bibliométricas, que se apresentam no atual contexto como instrumentos para mapear a atividade científica de um país, de uma área, de uma Instituição, ou mesmo de grupos de pesquisa. São parâmetros empregados como medidas indiretas da atividade de pesquisa científica e contribuem para compreensão dos objetivos da pesquisa, a estrutura da comunidade científica e do seu impacto social, político e econômico (KOBASHI, SANTOS, 2008).

O advento da tecnologia propiciou um grande desenvolvimento e acrescentou nova dimensão à análise da literatura científica. Atualmente, é possível processar

rápida e automaticamente grandes volumes de informação e obter indicadores variados. Nas instituições de pesquisa, os resultados de estudos bibliométricos fornecem elementos para avaliar e quantificar os níveis de produtividade, influenciando a distribuição dos recursos financeiros e a priorização das linhas de pesquisa, ou ainda para evidenciar o nível de excelência dessas instituições.

É nesse contexto que se situa o presente trabalho, desenvolvido em um instituto público de pesquisas criado em 1956 com a missão de desenvolver pesquisas e tecnologia na área nuclear e ciências relacionadas. O Instituto de Pesquisas Energéticas e Nucleares (IPEN) dispõe de uma expressiva produção científica gerenciada por sua unidade de informação por meio da manutenção de uma base de dados elaborada especialmente para essa finalidade. A citada base de dados adota padrões internacionais para realizar o processo de indexação dos trabalhos depositados. A produção científica provém principalmente das atividades de pesquisa do curso de pós-graduação, iniciado na Instituição em 1976. Dentro desse material, entre outros, encontram-se registradas e depositadas as dissertações e teses, objeto de estudo deste trabalho.

Sabe-se que a pesquisa é o fundamento sobre o qual se constrói uma ciência, e as dissertações e teses são documentos importantes para se estudar o estado da arte de determinada área de conhecimento. Pela experiência da autora no tratamento técnico da literatura científica e mais especificamente na indexação das dissertações e teses, observou-se que essas se constituem em uma das primeiras fontes de produção de ciência, validadas e certificadas, além de serem construídas dentro dos rigorosos padrões científicos. Dito de outra forma, a análise desse tipo de material resgata, rememora e avalia o conhecimento produzido pela Instituição ao longo dos 34 anos de existência do curso de Pós-graduação.

1.1 Objetivos

1.1.1 Geral

elaborar indicadores científicos a partir das dissertações e teses registradas na base de dados sobre a produção científica gerenciada pela unidade de informação do IPEN.

1.1.2 Específicos

- mapear numérica e tematicamente as dissertações e teses para identificar a trajetória, as frentes de pesquisa e as tendências desse tipo de literatura ao longo da história do programa de Pós-graduação do IPEN;
- verificar o índice de produtividade das teses no que diz respeito à publicação de artigos científicos, e o tempo transcorrido para sua publicação a partir da elaboração das teses; verificar, em nível macro, a qualidade dos canais de comunicação utilizados para publicação dos artigos;
- desenvolver e testar uma metodologia para identificar a correlação dos artigos publicados a partir das teses;
- testar a qualidade dos dados armazenados na base de dados e a viabilidade de utilização no processo de mensuração da produção científica.

1.2 Justificativa

O que primeiro justifica uma pesquisa é satisfazer uma curiosidade, é a necessidade de se conhecer mais e melhor o que se faz e como se faz.

Após longo período de trabalho com o controle, a sistematização, a disseminação e, principalmente, com a tarefa de fornecer dados atualizados e cada vez mais apurados sobre a produção científica para a Direção desta Instituição, a autora sentiu a necessidade de obter um conhecimento mais profundo da utilização da metodologia de mensuração, no processo de produção da ciência, bem como testar o potencial multifuncional das bases de dados construídas na Instituição. Isto conduziu à formulação de questões da seguinte natureza:

- é possível obter indicadores de produção científica a partir de dados empíricos, especificamente dos itens de literatura denominados dissertações e teses?
- as bases de dados que abrigam a produção científica de uma Instituição de pesquisa estão aptas a fornecer matéria-prima para a elaboração de indicadores?

Com base na experiência prática e na fundamentação teórica adquirida, formulou-se uma hipótese positiva, que somente por meio de um trabalho realizado sob os rigores científicos pode ser confirmada ou não.

Há 41 anos a documentação produzida na área nuclear no Brasil está indexada e controlada, devido ao trabalho do Centro de Informações Nucleares (CIN), como representante do *Internacional Nuclear Information System (INIS)* no Brasil. O CIN pertence à Comissão Nacional de Energia Nuclear (CNEN) no Rio de Janeiro e centraliza a coleta dessas publicações, mantendo um registro sistemático que cuida da disseminação dessa informação e a enviando para a base de dados especializada em Viena (Áustria). Apesar desse controle com a literatura gerada no país, não há registro da aplicação de técnicas bibliométricas para produzir indicadores baseados em produção científica da área nuclear e ciências relacionadas, seja em nível nacional ou regional, como no caso deste instituto, ou seja, após criteriosas pesquisas bibliográficas nas fontes de informação, não se localizaram estudos similares em nível nacional.

Com frequência, para se elaborarem trabalhos dessa natureza são utilizados os dados de bases de dados internacionais, que não são exaustivas, carregam um alto viés de representatividade e não registram esse tipo de material (as dissertações e teses). No entanto, as dissertações e teses são reflexos importantes da atividade de pesquisa institucional. Assim, o trabalho se justifica pela contribuição aos diversos segmentos.

A primeira contribuição que se vislumbra está direcionada aos gestores de política científica da Instituição.

Os indicadores bibliométricos não são um fim em si mesmos; eles são meios para fins específicos, contribuem com uma resposta para uma pergunta específica e servem de base para avaliações, recomendações ou ações futuras; conhecendo-se o histórico do desempenho passado, é possível planejar o futuro.

O contexto em que são produzidas as dissertações e teses está vinculado a um rigoroso processo de avaliação periódica. Há hoje oito programas de pós-graduação na área nuclear no país, dos quais sete estão concentrados na Região Sudeste e um na Região Nordeste, assim avaliados pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES¹).

- três cursos com nível 3 (Mestrado);
- dois cursos com nível 4 (Mestrado e Doutorado);
- um curso com nível 5 (Mestrado e Doutorado);
- dois cursos com nível 6 (Mestrado e Doutorado).

O programa do IPEN foi um dos primeiros a serem criados, na área, em nível regional (São Paulo); o reconhecimento e a qualidade do programa foram sendo construídos ao longo destes anos. Tão importante e árduo como avançar na qualificação estabelecida pela CAPES é manter o nível de reconhecimento e qualidade atribuído.

Desde 2001, o Programa está classificado, com nível 6 dentro da escala de 7 estabelecida pela CAPES. O curso conta atualmente com um número aproximado de 450 alunos e 130 orientadores credenciados; Nos últimos cinco anos produziu uma média de 123 dissertações e teses por ano. Em 2010, o curso de PG completou 34 anos de existência, e contabilizou 1.813 dissertações e teses defendidas.

¹ Trata-se de um processo de avaliação dos programas de pós-graduação desenvolvido pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), órgão do Ministério da Educação Brasileiro. Esse processo compreende a realização do acompanhamento anual e da avaliação trienal do desempenho de todos os programas e cursos que integram o Sistema Nacional de Pós-graduação, cujos resultados são expressos pela atribuição de uma nota na escala de "1 a 7". Site <http://www.capes.gov.br>

Portanto, considerou-se oportuno iniciar um estudo dessa natureza uma vez que, de acordo com VAN RAAN (2005), para estudos quantitativos desse tipo é necessária a existência de séries temporais de dados.

A segunda e não menos importante contribuição que se vislumbra diz respeito à prática profissional de gerenciamento da produção científica desenvolvida em unidades de informação especializadas; em outras palavras, objetiva-se propiciar uma reflexão das práticas aos profissionais da informação, área de atuação da autora.

Pretende-se demonstrar a utilidade e o potencial das bases de dados construídas em bibliotecas nacionais, particularmente nas bibliotecas especializadas e universitárias, gestoras, em sua grande maioria, da produção científica gerada nas suas respectivas instituições.

Integra a formação curricular do profissional da informação a construção de bases de dados. Esses repositórios de informação foram construídos inicialmente com a finalidade de armazenar, reunir, organizar documentos representados pelos metadados e possibilitar a busca e localização de informações referenciais, atualmente com forte tendência de mudança para informações textuais. No entanto, as bases de dados planejadas, estruturadas e devidamente codificadas se configuram como dispositivos multifuncionais com grande potencial de utilização na produção de indicadores para avaliar o estado da arte da ciência e tecnologia (IGAMI, 2007).

A aplicação de técnicas bibliométricas requer bases de dados organizadas e estruturadas de forma adequada para a produção de indicadores. A bibliometria se baseia sempre em documentos publicados e só se concretiza quando há uma documentação preparada adequadamente (IGAMI, 2007).

A ideia de agregar valor à informação armazenada nas bases de dados apoia-se no *continuum* do conhecimento, proposto por TAYLOR (1996), em seu livro *The value added information system*.

Figura 1 – O *continuum* do conhecimento. Modelo proposto por Taylor

Fonte: TAYLOR, R. *The value added information system* 1996 p.83

Nesse livro, TAYLOR² apresenta um modelo prático e conceitual de produção de informações com valor agregado mediante a premissa de criação de um processo capaz de transformar dados sem nenhum significado em informação útil, isto é conhecimento produtivo, para subsidiar as ações de gestão.

O modelo do valor agregado é de aplicação ampla, podendo ser utilizado tanto em bibliotecas como em qualquer unidade que gerencie a informação. Observa-se na FIG. 1 um caminho a ser percorrido pelas unidades de informação especializadas

² Robert Saxton Taylor, historiador, bibliotecário, autor de vários livros na área, defendia a ideia de um modelo de uso da informação centrado no usuário, foi presidente da *American Society for Information Science*, em 1992.

em que a maioria das bibliotecas no país atua até o nível 1. Porém, agregando valor a suas atividades, trabalhando a massa documental, é possível avançar no *continuum* do conhecimento e aperfeiçoar seu desempenho, percorrendo as etapas 2 e 3 para fortalecer sua imagem no contexto da Instituição mantenedora.

Acredita-se, também, que o trabalho possa contribuir para a consolidação dos estudos métricos no país. Os estudos que utilizam a bibliometria foram bastante empregados na década de 1970, permanecendo estagnados na década seguinte. A partir da década de 1990, o interesse pela área cresceu devido ao grande desenvolvimento da tecnologia da informação, à facilidade no manuseio automático de grandes volumes de dados e ao desenvolvimento de softwares para tratamento e visualização de dados.

Atualmente encontram-se estudiosos especializados no assunto em muitos países, estabeleceram-se grupos e redes de pesquisadores trabalhando no desenvolvimento de métodos e técnicas em bibliometria que VAN RAAN (1993) denomina como avançada, NOYONS e VAN RAAN (2002) (CWTS Leiden, Holanda), GLANZEL (2006) (Bélgica), GEISLER (2000) (EUA), entre muitos outros especialistas no assunto.

A exemplo do que ocorreu em nível internacional, no Brasil, hoje, a bibliometria se constitui em uma área em pleno desenvolvimento demonstrado pelo crescente número de trabalhos obtidos nas pesquisas bibliográficas efetuadas em revistas nacionais como *Perspectivas em Ciência da Informação*, *Ciência da Informação*, *Transinformação* bem como em participação em eventos nacionais e internacionais, onde são apresentados trabalhos utilizando as técnicas bibliométricas nos mais diversos *corpus* de estudo. Observam-se também núcleos de pesquisa dedicados ao estudo e à familiarização das métricas da ciência. Entre os núcleos de pesquisa cadastrados no diretório do CNPq citam-se alguns exemplos: indicadores de Ciência, Tecnologia e Inovação, Universidade Federal de São Carlos (UFSCar), Comunicação e Divulgação Científica do Instituto Brasileiro de Informação Científica e Tecnológica/Universidade Federal Fluminense (IBICT/UFF), Comunicação Científica, Universidade de Brasília (UnB), Comunicação Científica, Universidade Federal Rio Grande do Sul (UFRGS).

O que este trabalho não aborda?

Nesta oportunidade, não se buscou abordar as questões que envolvem o desempenho individual dos autores das dissertações e teses, como índices de citação dos autores, índice H ou índice de produtividade individual; também não se analisou o conteúdo ou a forma das teses, isto é, qualidade ou alinhamento com as áreas de concentração do Programa. O conteúdo das teses somente foi analisado pela abrangência geral e categorizado dentro das metacategorias temáticas estabelecidas pela *Subject Categories and Scope Descriptions do International Nuclear Information System (INIS)*³ (INTERNATIONAL ATOMIC ENERGY AGENCY , 2002).

1.3 Apresentação e organização deste trabalho

Trata-se de um estudo multidisciplinar que recorre a conhecimentos de pelo menos três áreas de conhecimento: estudos sociais da ciência (sociologia) e organização e representação do conhecimento (biblioteconomia) e estatística.

³ Base de dados internacional especializada em literatura nuclear e ciências relacionadas; o sistema INIS (www.iaea.org/inis) pertence à *International Atomic Energy Agency (IAEA)* com sede em Viena (www.iaea.org). O sistema INIS está estabelecido há 42 anos e a manutenção da base é realizada de forma descentralizada, isto, é todos os países membros da IAEA contribuem com a literatura nuclear do país. Devido a este fato o INIS emite manuais, normas, procedimentos para a padronização de alimentação da base, elabora e distribui um Tesouro especializado e a lista de metacategorias para orientação dos indexadores e catalogadores da base, bem como, periodicamente, promove cursos de capacitação para indexadores em Viena. No Brasil, o representante do INIS que coordena toda a coleta e envia a literatura nuclear é o Centro de Informações Nucleares (CIN) da Comissão Nacional de Energia Nuclear no Rio de Janeiro (INTERNATIONAL NUCLEAR INFORMATION SYSTEM, 2010).

O trabalho segue a estrutura para a elaboração das dissertações e teses vigente na Instituição, aprovada pela Comissão do Programa de Pós-graduação do IPEN em Tecnologia Nuclear, e está estruturado em 9 partes⁴ (FIG. 2):

Figura 2 – Estrutura adotada para a elaboração deste trabalho

⁴ A denominação “Partes” só foi empregada para descrever a estrutura do trabalho neste quadro.

2 O INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES, IPEN-CNEN/SP

O local onde se desenvolveu este trabalho é o Instituto de Pesquisas Energéticas e Nucleares (IPEN). Trata-se de uma autarquia pública gerida administrativamente pela Comissão Nacional de Energia Nuclear, com sede no Rio de Janeiro, uma Instituição federal subordinada ao Ministério de Ciência e Tecnologia. Foi criado com o nome de Instituto de Energia Atômica, fruto de um convênio celebrado entre o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Universidade de São Paulo, pelo Decreto federal n. 39.872, de 31 de agosto de 1956 (BRASIL, 1956).. Integrava este convênio a construção de um reator experimental.

No mesmo ano foram iniciadas as obras do prédio que abrigaria o primeiro reator nuclear de pesquisas do hemisfério sul, O *IEA-R1*, que foi doado pelo governo norte-americano dentro do Programa *Átomos para a Paz*. Desde os primeiros anos, o reator do IPEN foi utilizado para a formação de especialistas, desenvolvia pesquisas nas áreas de metalurgia, física nuclear e eletrônica, bem como pesquisas sobre elementos combustíveis que viriam a ser utilizados no próprio reator *IEA-R1*. Uma década mais tarde o Instituto já adquiria projeção internacional em pesquisas nucleares, produção de radiofármacos e substâncias marcadas para uso na medicina nuclear (ANDRADE, 2006).

O IEA passou a integrar a estrutura da Comissão Nacional de Energia Nuclear em 1963, onde permaneceu até 1972, quando se tornou uma autarquia estadual associada à USP. A integração com a Universidade sempre existiu, uma vez que os primeiros pesquisadores pertenciam ao quadro docente da USP (ANDRADE, 2006).

Em 1979, a Instituição mudou de nome objetivando ampliar o escopo de suas atividades, passando a denominar-se Instituto de Pesquisas Energéticas e Nucleares, IPEN, vinculado à Secretaria de Ciência e Tecnologia do Governo do Estado de São Paulo. Em novembro de 1982, o Governo do Estado de São Paulo firmou convênio com a Comissão Nacional de Energia Nuclear (CNEN), órgão

federal pelo qual passou a ser gerido técnica e administrativamente; mudou a denominação para IPEN-CNEN/SP, e assim permanece até os dias atuais (ANDRADE, 2006).

O IPEN esta localizado no campus da Cidade Universitária (USP), São Paulo, ocupando uma área de 500.000 m². Somando seus funcionários, hoje cerca de 1.000, alunos que frequentam o curso de Pós-graduação, as disciplinas optativas, além de colaboradores, fornecedores, terceirizados, funcionários do Centro de empresas incubadas (CIETEC) e ainda o pessoal militar e civil do Centro Tecnológico da Marinha (CTM/SP), circulam no campus do IPEN diariamente mais de 2.500 pessoas (INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES, 2007-2010).

Em sua trajetória, o IPEN tem se destacado nas pesquisas e aplicações das técnicas nucleares, na produção de radiofármacos para uso na medicina nuclear em todo o país, fontes radioativas, ligas metálicas e cerâmicas, materiais utilizados no setor nuclear, entre outros. Além das atividades de pesquisa, produção de radiofármacos e serviços especializados que utilizam as técnicas nucleares, o IPEN tem-se destacado pela difusão dos conhecimentos por meio do ensino efetuado pelo Programa de Pós-Graduação. Devido ao fato de a Instituição se encontrar sob a administração Federal, as mudanças nos Programas de Governo, principalmente do ponto de vista de financiamento da atividade de pesquisa, têm reflexos imediatos para a Instituição (INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES, 2007-2010).

2.1 A área nuclear no Brasil

Ao se abordar o tema da área nuclear, bastante polêmico no país e no mundo, não se pretende abordar em profundidade a atuação, o histórico ou o aspecto político e técnico da área, por se entender não ser este o foco deste trabalho. No entanto, faz-se necessário um breve relato histórico para adiante se entender como essas questões interferiram na produção de ciência dentro da instituição IPEN.

Historicamente o desenvolvimento da tecnologia nuclear sempre esteve relacionado a questões estratégicas como, por exemplo, o desenvolvimento de armamentos nucleares; trata-se de uma tecnologia sensível, ou seja, que permite o uso civil, mas com possibilidades de emprego militar (MARQUES, 1992). Entretanto, a energia nuclear tem aplicações práticas e a principal está na geração de energia elétrica. Muito pouco se divulga sobre os benefícios do bom uso da energia nuclear, dos resultados e contribuições práticas das pesquisas no dia a dia das pessoas. A área médica, alimentícia, o setor produtivo, industrial, o meio ambiente e outros, além da citada geração de energia, são exemplos da grande utilização e aplicação das técnicas nucleares (LOUREIRO, 1980).

A década de 1940 foi marcada pela grande conflagração mundial iniciada em 1939, e configura-se como um período conturbado tanto política, econômica como socialmente, porém notável quanto ao desenvolvimento da Ciência e Tecnologia, com o surgimento do que se denominaram novas tecnologias, entre elas especialmente a área de energia nuclear, revelada ao mundo mediante a detonação das bombas atômicas no Japão, desenvolvidas no Projeto Manhattan, um programa de pesquisas sem precedentes na história mundial (MARQUES, 1992).

O impacto dessas novas tecnologias foi sentido nos diversos segmentos sociais do país, sobretudo o da energia nuclear, que, pelo seu potencial, obviamente despertou interesse tanto entre os pesquisadores como entre os militares. Para os pesquisadores era uma prova irrefutável da importância da pesquisa científica, a energia nuclear era um argumento que poderia ser utilizado para a implementação de um verdadeiro sistema de desenvolvimento em Ciência e Tecnologia. Foi nesta época que foi criado o Conselho Nacional de Pesquisas (CNPq) com a perspectiva de estabelecer todo um sistema de Ciência e Tecnologia no país e de manter o controle da industrialização e exportação dos materiais físséis (ANDRADE, 2006).

O fato de o Brasil possuir reservas significativas de material físsil, principalmente tório, é um fato relevante, pois em virtude desta riqueza o Brasil assinou um acordo com os Estados Unidos (EUA) em 1945, pelo qual se comprometia a vender com exclusividade para aquele país 5 mil toneladas anuais de

monazita, matéria-prima importante para a fabricação do combustível nuclear ANDRADE (2006); MARQUES (1992).

MARQUES (1992), adota uma divisão temporal de três períodos distintos para entender a história da política nuclear brasileira, são eles: a fase nacionalista (1949-1954), a fase diplomática (1955-1974) e a fase do desenvolvimento dependente que se inicia em 1975, que ele entende estender-se até os dias atuais, sendo a mais marcante e longa.

Na fase nacionalista, da década de 1940, e mais especificamente ao fim da Segunda Guerra Mundial, os Estados Unidos propuseram nas Nações Unidas um tratado internacional que criaria uma autoridade mundial responsável pela gestão de todas as reservas de urânio do mundo, o chamado Plano Baruch. O Brasil, representado pelo almirante Álvaro Alberto da Mota e Silva, juntamente com a outrora União Soviética foram os únicos países a se oporem a este plano por entenderem que ele assegurava aos Estados Unidos o monopólio da tecnologia e da matéria-prima para a fabricação do combustível nuclear. Nesta ocasião o representante brasileiro formulou pela primeira vez o *princípio das compensações específicas*; por meio deste princípio o Brasil se comprometia a fornecer matéria-prima (monazita e terras raras, abundante no país) para os países interessados, desde que estes fornecessem em troca tecnologia e equipamentos para o desenvolvimento de uma indústria nuclear brasileira. O almirante configurava-se assim como pioneiro na formulação de uma política nuclear independente para o país (MARQUES, 1992).

Em 15 de janeiro de 1951, foi criado o Conselho Nacional de Pesquisas (CNPq). Entre as atribuições do CNPq estavam *a investigação e a industrialização da energia atômica e suas aplicações*. Presidido pelo almirante Álvaro Alberto, o CNPq logo implementou a política das *compensações específicas*, angariando com isto forte oposição dos Estados Unidos, não conseguindo sucesso na obtenção da transferência da tecnologia nuclear. O almirante, então, volta suas negociações para a Alemanha, conseguindo, em 1953, três unidades de enriquecimento de urânio pelo processo de ultracentrifugação. Os conjuntos foram interceptados por um alto

comissário norte-americano, cumprindo ordens do presidente da Comissão de Energia Atômica dos Estados Unidos. A possibilidade de o Brasil desenvolver a tecnologia para enriquecimento do urânio desagradava aos Estados Unidos. Após inúmeras gestões diplomáticas do almirante Álvaro Alberto, os equipamentos foram liberados e chegaram ao país LOUREIRO (1980); ANDRADE (1996); MARQUES, (1992).

Conforme comenta MARQUES (1992), a fase nacionalista foi caracterizada pela luta nacionalista liderada pelo almirante Álvaro Alberto, e nesta fase já havia o interesse de promover a transferência de tecnologia, mostrando que os próprios dirigentes brasileiros conheciam com clareza o que significava dominar uma tecnologia de ponta e sua estreita relação com a pesquisa, contudo não foi possível prover a pretendida capacitação nuclear, sendo os principais obstáculos o obsolescência do parque industrial brasileiro daquela época, a dificuldade de uma política de aquisição de maquinaria para modernizá-lo, aliados à sabotagem e às pressões americanas de impedir avanços próprios nas pesquisas nesta área.

Com a morte do presidente Getúlio Vargas e a exoneração do almirante Álvaro Alberto da presidência do CNPQ, termina a chamada fase nacionalista (MARQUES, 1992).

Em 1955, o presidente Café Filho assina a integração do Brasil no programa americano *Átomos para a Paz*. Os Estados Unidos sinalizavam assim que haviam desistido de impedir o acesso de outros países às tecnologias nucleares, e em vez disso procuravam mantê-los sob seu controle e vigilância. Na fase diplomática, é o Itamaraty que vai dominar e conduzir as negociações. Neste mesmo ano de 1955 foram assinados acordos com os Estados Unidos, acordo de cooperação para o desenvolvimento da energia atômica com fins pacíficos, e o programa conjunto para o reconhecimento e a pesquisa de urânio no Brasil, este previa a pesquisa e a avaliação das reservas de urânio que seriam vendidas para os Estados Unidos (LOUREIRO, 1980).

Em 1956, com a posse do presidente Juscelino Kubitschek, iniciou-se não uma nova fase, porém, uma nova conjuntura para a política nuclear brasileira,

notadamente uma tentativa de resgate da autonomia do setor. São criados neste mesmo ano o Instituto de Energia Atômica em São Paulo (IEA), atualmente o Instituto de Pesquisas Energéticas e Nucleares (IPEN), e iniciada a instalação e construção do primeiro reator de pesquisas do hemisfério sul, o IEA-R1, de origem americana, e a Comissão Nacional de Energia Nuclear (CNEN) no Rio de Janeiro, com a responsabilidade de propor as medidas necessárias à orientação de uma política de energia atômica. Em 1963 foi criado o Instituto de Engenharia Nuclear, em um convênio da CNEN com a Universidade Federal do Rio de Janeiro (UFRJ) LOUREIRO (1980); ANDRADE (2006); MARQUES (1992).

O Brasil já contava com físicos nucleares de renome internacional, como Cesar Lattes, Mario Schonberg, Marcelo Damy de Souza Santos, José Leite Lopes e outros. Foi um período importante na capacitação de recursos humanos, e foi em torno destes institutos que avançaram os trabalhos de pesquisa procurando identificar soluções para o desenvolvimento da tecnologia nuclear, diante das barreiras impostas para a obtenção externa desta tecnologia (ANDRADE, 2006).

O Brasil viria a se envolver em problemas diplomáticos com os Estados Unidos, devido à recusa brasileira de não adesão ao Tratado de Não Proliferação de Armas Nucleares, aprovado pela Organização das Nações Unidas (ONU), em 12 de junho de 1968. A política nuclear brasileira sempre se pautou pelos objetivos de usos pacíficos da energia nuclear, o Brasil considerou que *o tratado não estabelecia um balanço aceitável entre os direitos e as obrigações das nações nucleares e não nucleares*, isto é, o país alegou que o tratado era um limitador da soberania nacional (MARQUES, 1992).

Em 1970 a CNEN abre uma concorrência internacional para a compra do reator de Angra 1, na cidade de Angra dos Reis, Rio de Janeiro, que foi vencida pela empresa norte-americana Westinghouse, subsidiária da General Electric. Nesse contrato ficou acertado que o reator de potência vendido ao Brasil seria um reator de água pressurizada (PWR). Tratava-se de um contrato em que a última coisa que ocorreria seria a transferência de tecnologia. Em 1972 o Brasil assinou um novo acordo com os Estados Unidos, por meio do qual aquele país forneceria urânio

enriquecido em troca de urânio natural brasileiro. Pelo acordo também a Comissão Nacional de Energia Atômica dos Estados Unidos assumiria o controle das instalações brasileiras para que estas não fizessem uso militar. Outros países em desenvolvimento também trabalhavam em projetos de reatores de potência como a Argentina e a Índia. Em 1974 a Índia detona sua primeira bomba atômica, fato que dificultou significativamente as negociações com os americanos para a construção do reator brasileiro.

Em virtude destas dificuldades de transferência de tecnologia, as políticas restritivas ao fornecimento de urânio enriquecido somadas à crise do petróleo provocada em 1973 pelos produtores árabes ao decidiram impor um significativo aumento nos preços dessa matéria-prima repercutiram negativamente, fazendo com que o país reorientasse sua política nuclear e buscasse novos interlocutores, aproximando-se da Alemanha e assinando o acordo Brasil-República Federal da Alemanha, para fornecimento de tecnologia nuclear. Encerra-se assim a fase diplomática (MARQUES, 1992).

De acordo com MARQUES (1992), os acordos não produziram os resultados esperados porque nesta fase evidencia-se que as negociações foram canalizadas pelas vias diplomáticas, cujos integrantes desconheciam o significado do que é pesquisa, e muito menos o que é transferência de tecnologia, afastando com isso interlocutores da comunidade científica com maior capacidade para negociação de questões técnicas e complexas. LOUREIRO (1980), comenta que no campo nuclear não bastaria ao Brasil dispor de reatores, era necessário sobretudo dominar o ciclo do combustível nuclear para evitar crises como a do petróleo. Os técnicos da CNEN e da Companhia Brasileira de Tecnologia Nuclear (CBTN) já estudavam meios de dominar toda a tecnologia do ciclo do combustível⁵, e os institutos de pesquisa seriam os locais apropriados para esta finalidade, porém, a experiência de outros países mais industrializados mostrava que o desenvolvimento de um processo próprio desta natureza era longo, em torno de 15 a 20 anos. Além do que, a

⁵ O ciclo do combustível foi desenvolvido por pesquisadores no IPEN/CNEN-SP e deu origem a um número significativo de dissertações e teses, na categoria temática de Química, conforme pode ser observado neste trabalho na p. 112, FIG. 17.

execução de um programa desta complexidade seria impossível sem a disponibilidade de recursos humanos capacitados.

MARQUES (1992), confirma que ficou claro que não era possível lograr êxito no domínio de uma tecnologia tão complexa sem um ambiente propício para a prática da pesquisa científica e tecnológica, ou seja, sem a participação de universidades e institutos de pesquisa, onde se desenvolvessem substratos mentais e técnicos de alta qualidade.

Percebendo isso, a CNEN, o MEC e o CNPq realizaram estudos sobre as necessidades de mão de obra especializada; as conclusões do estudo levaram o presidente Geisel a assinar em julho de 1976 um decreto criando o Programa de Recursos Humanos para o setor nuclear, o PRONUCLEAR, previsto para durar 10 anos e formar, entre 1976 e 1986, 4.335 profissionais de nível superior e 5.580 de nível médio (LOUREIRO, 1980). Em 1983, o programa foi desacelerado, e finalmente desativado em 1986 (PRONUCLEAR, 1986)⁶.

O acordo Brasil-Alemanha incluía atividades abrangentes, como é o caso do desenvolvimento do ciclo do combustível nuclear, que inclui etapas que vão desde a prospecção de minérios atômicos até o enriquecimento do urânio e o reprocessamento do plutônio, materiais considerados potencialmente bélicos, motivo pelo qual se intensificaram as desconfianças, por parte da sociedade e da comunidade científica, de que os militares procuravam condições para produzir artefatos bélicos.

Em setembro de 1987, o presidente Sarney anunciou o domínio do enriquecimento de urânio, e dava a conhecer a existência de um Programa Nuclear Paralelo desenvolvido pela Marinha nas instalações do IPEN em São Paulo, porém com o objetivo de desenvolver um protótipo de submarino nuclear (MARQUES, 1992).

Este Programa foi mantido por um longo período fora do conhecimento público; no ano seguinte ao anúncio, a empresa Centrais Elétricas Nucleares S/A,

⁶ No IPEN/CNEN-SP foram formados 131 mestres dentro do PRONUCLEAR e 35 bolsistas foram enviados ao exterior para realizar o doutorado, ver p.139 deste trabalho para mais detalhes.

(NUCLEBRÁS) foi extinta e o programa nuclear paralelo foi incorporado às pesquisas oficiais, mais exatamente, o Decreto-Lei n. 2.464, de 31 de agosto de 1988 (BRASIL, 1988), oficializou o programa paralelo e trazido ao conhecimento da sociedade brasileira.

Pelo Decreto também ficou determinado o término da construção de Angra II e III após reavaliação do acordo com a Alemanha bem como a construção de um reator em Iperó, São Paulo, para servir de protótipo para os reatores do submarino nuclear (KURAMOTO, 2002). O controle das obras das centrais nucleares Angra 2 e 3 é atualmente gerenciado pela Eletronuclear, uma estatal ligada à CNEN e que administra as atividades de Angra ainda obedecendo aos acordos firmados com a Alemanha (ENERGIA NUCLEAR, 2000).

Um fato significativo também marca negativamente a história da energia nuclear no país. Em 17 de setembro de 1990, em um gesto altamente político, o presidente Fernando Collor fechou a área de teste na Serra do Cachimbo, um fosso profundo que havia sido preparado pelos militares para realizar provas subterrâneas, e uma semana depois anunciou no Conselho das nações Unidas que o Brasil rejeitava a ideia de qualquer teste que implicasse explosões nucleares mesmo com fins pacíficos, encerrando a faceta militar do programa de experiências nucleares (KURAMOTO, 2002).

BUSTANI⁷, em 2005, em entrevista para a revista *Pensar Brasil*, analisa o programa nuclear sob as perspectivas jurídica, econômica, ambiental, política e estratégica. Afirma que do ponto de vista jurídico não existe impedimento para que o Brasil faça uso da energia nuclear com fins pacíficos. O país já é signatário do Tratado de Não Proliferação Nuclear (TNP)⁸, além do que todas as instalações

⁷ Jose Mauricio Bustani, diplomata de carreira e embaixador do Brasil em Londres. Foi diretor da Organização para Proibição de Armas Químicas (OPAQ) de 1997 a 2002.

⁸ Criado em 1997, o documento autoriza a IAEA a inspecionar as instalações previstas pelo TNP e qualquer outro lugar do país considerado passível de ser utilizado em um programa nuclear, como centros de pesquisa ou usinas. O Protocolo também prevê o amplo acesso da Agência à tecnologia nuclear utilizada no país, inclusive aos detalhes dos sistemas de enriquecimento de urânio (FERREIRA FILHO 2010)

nucleares estão sujeitas a inspeções por parte da *International Atomic Energy Agency* (IAEA).

Na Constituição brasileira consta o veto para o uso da energia nuclear com fins não pacíficos, e a CNEN mantém um elevado padrão de fiscalização.

Do ponto de vista econômico é visível que o país com a 5º maior população do mundo precisará de energia para manter e ampliar seu desenvolvimento e reduzir a dependência externa com relação à importação de petróleo. Embora a energia nucleoeletrica seja mais cara do que a hidroeletrica, já há sinais perceptíveis de que o uso desta última se aproxima dos limites, sendo oportuno um planejamento em longo prazo para diversificar a matriz energética.

Do ponto de vista ambiental, a energia nuclear é a que mais se apresenta como uma opção viável e a fonte mais limpa. O reconhecido ambientalista James Lovelock e Patrick Moore, fundador do *Greenpeace*, mudaram sua opinião e passaram a defender a opção energia nuclear, por entenderem que se trata na verdade de uma fonte energética viável, não poluente e com boa relação custo/benefício BUSTANI (2005); AGUIAR (2008).

BUSTANI (2005) analisa que da perspectiva política, a energia nuclear continua a se constituir em uma energia sensível, pelo seu potencial bélico, porém a adesão ao TNP, no qual o Brasil renuncia à fabricação de armas nucleares, garante que não há desvios das atividades pacíficas, e um meio de manter esta credibilidade é implementar medidas de controle apropriadas por meio de organismos internacionais, para aprimorar as condições de segurança nacional.

Da perspectiva estratégica, BUSTANI (2005), entende que o país não deve renunciar ao progresso do conhecimento nuclear, pois este tem aplicações promissoras em campos diversos como a medicina nuclear, a agricultura e outras áreas já em desenvolvimento por meio dos institutos de pesquisa atuantes no país. Em suma, o país não deve abdicar do uso e do desenvolvimento de tecnologias que possam contribuir pacificamente para um projeto nacional de prosperidade.

Como pode ser observado pelo breve histórico, a área nuclear desde o início da sua utilização, na década de 1950, passou por diversas fases favoráveis ou

menos favoráveis, mas foi na década de 1990, durante o governo Collor, por questões políticas, que aconteceu a mais severa desaceleração da área. Ações como essa repercutiram fortemente no desenvolvimento da área em todo o país.

Em 2005, após duas décadas sem planejamento, a área nuclear voltou a ter um programa nuclear com metas e diretrizes claras e um planejamento em longo prazo. Foi criado um grupo ministerial, constituído pelo presidente da República e liderado pelo Ministério de Ciência e Tecnologia (MCT), para estudar e apresentar o Novo Programa Nuclear Brasileiro (PNB) com um planejamento a se estender até 2030 (ARANHA, 2005).

Entre as recomendações apresentadas pelo grupo, consta em primeiro lugar que a geração nuclear permaneça no país como um dos componentes da matriz energética, e o governo, como visão de futuro para o Programa Nuclear Brasileiro, em um cenário mais avançado, recomenda que a participação da energia nuclear na malha energética seja elevada para 4,2% ou 5,7%. O planejamento envolve a participação de todos os institutos de pesquisa no país, cada um com sua área de excelência (ARANHA, 2005).

Nos últimos anos, dois fatos importantes reiniciaram os debates na sociedade brasileira sobre a utilização da energia nuclear. Internacionalmente foi publicado em 2007:

- a)** o relatório do Painel Intergovernamental sobre Mudanças Climáticas (IPC⁹ – sigla em inglês), o qual confirmou que o aquecimento global é provocado pelo descontrole de emissões de gases produzidos pela queima de combustíveis fósseis, e entre outras medidas, recomenda o uso da energia nuclear como uma alternativa climática;
- b)** o segundo fato importante aconteceu em junho de 2007, quando o Conselho Nacional de Política Energética (CNPE) aprovou o Plano Nacional de Energia (PNE) 2030, que traça metas de investimento para o setor para os próximos

⁹ *Intergovernmental Panel on Climate Change-IPCC* Instituição internacional criada pela *United Nations Environment Program (UNEP)* e pela *World Meteorological Organization (WMO)* para avaliação das mudanças climáticas e o impacto socioeconômico destas em nível mundial (ORGANIZAÇÃO MUNDIAL DE METEOROLOGIA, 2007).

20 anos. Nesse cenário a energia nuclear assume papel mais importante e um crescimento progressivo.

A retomada do Programa Nuclear Brasileiro (PNB) deve proporcionar grande impulso na renovação dos quadros técnicos do setor; sabe-se que recursos humanos qualificados e experientes compõem a base estrutural do Programa, porém duas décadas sem grandes investimentos na área já mostraram seus efeitos, criou-se o chamado “vácuo do conhecimento” como consequência da falta de renovação do pessoal e da aposentadoria daqueles que estavam altamente capacitados, e que não puderam transmitir seu conhecimento para os novos profissionais (XAVIER, 2009).

Ainda de acordo com o mesmo autor é preciso modelar uma nova política de formação de recursos humanos para o setor nuclear que seja compatível com o crescimento do Programa Nuclear Brasileiro (PNB), e esse planejamento inclui os programas de pós-graduação.

Pelo panorama resumido sobre a área nuclear, deduz-se que há muito trabalho a ser desenvolvido tanto em pesquisa como na formação de recursos humanos, bem como na aplicação de tecnologia nuclear nos próximos períodos.

Cabe refletir quanto do conhecimento acumulado, ao longo da trajetória do IPEN, poderia contribuir para o PNB? Certamente, são os gestores da política científica que podem responder a essas questões.

A maioria dos gestores que atuam no IPEN conhece e acompanha a atuação institucional. Porém para administrar com qualidade é necessário mais do que percepção; dados quantitativos, diagnósticos, cenários atualizados e tendências fundamentam e norteiam as decisões políticas, gerenciais, científicas a serem tomadas.

Em síntese, a área nuclear é um setor estratégico e importante para o país, porém demanda o uso de altíssima tecnologia, altos investimentos, recursos humanos capacitados, além de ser uma área fortemente sensível às críticas da sociedade, em grande parte devido ao desconhecimento do seu potencial. Ora a área nuclear é uma área estratégica indispensável para o desenvolvimento do país como alternativa para a produção e geração de energia elétrica, ora se trata de uma

área de alto risco, com tecnologia dispendiosa, o que fica claro depois deste breve histórico é que a política para a área nuclear no país nunca foi consistente, não se estabeleceu uma política de estado e sim de governo.

A partir de 2005 a área vivencia uma fase positiva, com a decisão política de incluir a Energia Nuclear na matriz energética nacional, a retomada da construção das usinas nucleares (ARANHA, 2005), e a concretização do projeto de Política Nacional de Energia Nuclear que propiciará, pela primeira vez no país, que o governo normatize o que se pretende fazer nesta área (KURAMOTO, 2002).

2.1.1 A institucionalização da área nuclear e o Programa de Pós-graduação

Identificar o perfil da ciência produzida por meio das dissertações e teses pode auxiliar não só no planejamento da política institucional, como também no conhecimento do grau de institucionalização da ciência nuclear e das áreas relacionadas dentro do Programa em questão.

BUNGE (1980), afirma que a ciência precisa de um arcabouço teórico, metodológico e temático para orientar as atividades de pesquisa, uma vez que este arcabouço representa os fundamentos e os limites territoriais e temporais da ciência, possibilitando sua institucionalização.

De acordo com GOLDFARB e FERRAZ (2002), são necessários quatro elementos para que uma disciplina se consolide como científica: ensino, pesquisa, divulgação e aplicação do conhecimento; de todos esses elementos, a pesquisa desempenha papel fundamental, sendo os outros elementos consequentes.

WHITLEY (1974), em seus estudos sobre a sociologia da ciência, destaca que a pesquisa representa a descoberta, a atualização e a evolução do conhecimento; uma ciência se oficializa pelo reconhecimento que sua própria comunidade e outras comunidades lhe atribuem. Destaca ainda que a ciência pode ser visualizada por duas vertentes: a *cognitiva* e a *social*.

A primeira responde pela clareza dos componentes do sistema conceitual e a segunda diz respeito à organização das estruturas formais e informais que legitimam

as atividades de pesquisa. Dito de outra forma, as questões teóricas, epistemológicas, interdisciplinares, os problemas abordados, a aceitabilidade das soluções, os métodos e instrumentos são domínio da institucionalização cognitiva; já a social se refere às estruturas formais de que uma área de conhecimento necessita para se tornar visível.

Quais são essas estruturas formais? WHITLEY (1974) as define como o grau de organização interna e externa, os instrumentos de divulgação, a estrutura acadêmica e a política institucional. Assim, os programas de pós-graduação são estruturas formais altamente institucionalizadas e sistematicamente avaliadas, bem como as dissertações e teses produzidas nesses programas se constituem em um dos principais instrumentos de divulgação da institucionalização social da ciência.

O Programa de Pós-graduação foi criado no então Instituto de Energia Atômica (IEA) em 1976. A gestão e responsabilidade ficaram a cargo do Instituto em associação com a Universidade de São Paulo. O Programa mostrou resultados positivos desde o início, sendo reconhecido pela CAPES como nível “A” a partir do início do sistema de avaliação do órgão, ou seja, tornou-se um paradigma de eficiência na formação de recursos humanos na área (ROGERO, 2009)¹⁰.

O modelo administrativo adotado, qual seja, um programa de pós-graduação gerido por um instituto de pesquisas em associação com uma universidade e conduzido por uma comissão de pós-graduação autônoma, tem mostrado resultados altamente satisfatórios, sendo imitado no país por outras instituições congêneres.

Em 1997, houve modificações nas áreas de concentração visando a uma melhor adequação, passando das duas áreas existentes: **1)** reatores nucleares de potência e tecnologia do combustível; e **2)** tecnologia nuclear básica, para três: **a)** aplicações; **b)** materiais; **c)** reatores nucleares. O Programa de Pós-graduação opera atualmente com estas três áreas de concentração.

¹⁰ Comunicação pessoal fornecida por ROGERO, J. R., em São Paulo, 2009.

3 QUADRO TEÓRICO

Para se compreender como se processa a construção da ciência é oportuno relembrar algumas conceituações sobre ciência, embora se trate de um conceito tão abstrato, amplo e complexo sobre o qual muitos estudiosos já se debruçaram ao longo da história da humanidade. FREIRE-MAIA (1998), diz que raramente os filósofos se propõem conceituar ciência e cita pelo menos três motivos: primeiro, porque se corre o risco de toda definição ficar incompleta, segundo, pela própria complexidade do tema e o terceiro, justamente pela falta de acordo entre as duas primeiras definições. Nesse sentido, MERTON (1979, p. 38), afirma que: *é quase unânime a noção de que a ciência é um vocábulo enganosamente amplo que abrange uma diversidade de coisas diversas embora relacionadas entre si.*

Segundo MORAIS (1988), a ciência é mais do que uma instituição, por ser um conceito abstrato, o que conhecemos de concreto são os cientistas e o resultado do seu trabalho. *O cientista contemporâneo sabe que não há nada de definitivo que tenha sido assentado por homens.* SCHWARTZMAN (1984), afirma que:

..."ciência" não é uma coisa simples, que se possa definir com facilidade recorrendo a uma boa enciclopédia. Trata-se de um fenômeno social e humano bastante complexo e variado, suficientemente importante para gerar todo um esforço para compreendê-lo e poder em seguida agir sobre ele. Esta é a origem da "ciência da ciência", e mais especificamente da sociologia da ciência, que trata de examinar o fenômeno científico como um fato social (SCHWARTZMAN 1984, p.16).

E com relação à atividade científica, SCHWARTZMAN (1984), comenta:

A atividade científica não é uma simples decorrência de características muito gerais do sistema econômico e social, mas depende de estruturas e sistemas sociais muito mais delicados e específicos. O trabalho científico exige grupos de pessoas dedicadas profissionalmente a ele; uma ética que valorize o conhecimento e prestigie aqueles que o busquem; um sistema de incentivos para o trabalho científico que lhe permita atrair os melhores talentos, e uma cultura que dê lugar ao surgimento de novos conhecimentos pela observação e a análise racional, em contraste com aquelas onde predominam os conhecimentos ritualizados e carregados de afetividade. O trabalho científico necessita,

ainda, que os cientistas sejam os principais avaliadores e juízes de seu trabalho, e que não tenham que submeter suas conclusões à aprovação de outras instâncias, religiosas, políticas ou institucionais (SCHWARTZMAN 1984, p.17).

Pelas conceituações infere-se que a ciência como atividade é eminentemente social, e independentemente do conceito que se adote, fica claro que a ciência avança por meio das atividades de pesquisa, e os resultados da pesquisa se materializam por meio de vários produtos, entre eles a produção científica. Isto é, a produção do conhecimento científico está fortemente associada à pesquisa, a qual segue princípios metodológicos para que a validade de seus resultados seja assegurada conforme afirma ZIMAN (1979).

Para BORDIEU (2004), a produção do conhecimento científico é um processo coletivo e dinâmico, é o universo do qual participam os agentes e as instituições que produzem este conhecimento; esse processo científico a que ele se refere tem uma estrutura própria que determina as ações que podem ou não ser realizadas. Esta estrutura é determinada pela distribuição do capital científico entre os agentes participantes desse processo científico.

BORDIEU (2004), também comenta que o capital científico é formado, principalmente, pelas publicações que despertam o interesse dos membros do campo científico e que retroalimentam a produção do conhecimento em um ciclo contínuo, com referência a esta dinâmica.

A este respeito LEVY (2001) afirma que os pesquisadores constroem o conhecimento científico a partir dos conhecimentos já existentes, citam trabalhos dos seus pares, inserem-se em um coletivo a qual eles pertencem o que pode ser interpretado como: a ciência se alimenta da própria ciência.

Para BUNGE (1980), o conjunto de conhecimentos de uma disciplina está em constante reformulação e evolução, o qual oferece um leque de possibilidades para novas pesquisas. As possibilidades advêm de tópicos convergentes, divergentes, deficientes, polêmicos e transversais, dos quais se originam novos campos de pesquisa

BRAGA (1974), registra que a literatura científica originou-se da correspondência trocada entre personalidades científicas da então incipiente Royal Society of London¹¹.

Porém, pode-se recuar ainda mais, uma vez que Arquimedes (inventor, físico, matemático) e Ptolomeu (matemático e astrólogo), a exemplo de outros autores clássicos da literatura, registraram seus tratados em forma de cartas aos amigos. Esse hábito perdurou por toda a Idade Média; o trabalho de Peter Peregrinus, que escreveu o primeiro tratado sobre os ímãs, por exemplo, foi enviado em forma de carta a um colega de Paris (essa carta foi considerada o primeiro artigo científico da história científica). No século XVIII o periódico já estava bastante difundido e no século XIX o artigo se tornou uma instituição e se situa no centro da comunicação científica (BRAGA, 1974).

A produção científica é vista, então, como a parte visível da atividade científica. As definições de produção científica variam de autor para autor, foram selecionadas aquelas que mais exemplificam este trabalho:

...Produção científica é a forma pela qual a universidade ou instituições de pesquisa se faz presente no âmbito da ciência; é a base para o desenvolvimento e a superação da dependência entre países e entre regiões de um mesmo país; é o vínculo para a melhoria da qualidade de vida dos habitantes de um país; é a forma de se fazer presente não só hoje, mas também amanhã; este rol pode ir longe mas, seja qual for o ângulo que se tome por referência, é inegável o papel da ciência na vida das pessoas, das instituições dos países. Pode-se afirmar que alguma produção científica está ligada à maioria ou quase totalidade das coisas, eventos, dos lazeres com que as pessoas se envolvem no cotidiano... (WITTER, 1997, p. 8).

LOURENÇO (1997) define produção científica como toda prova documental, independentemente do suporte, papel, meio magnético ou eletrônico, sobre determinado assunto de interesse de uma comunidade científica específica que

¹¹ Em 1662, formou-se a Royal Society of London, assim denominada porque o Rei Carlos II da Inglaterra concordou em conceder-lhe seu patrocínio. A Royal Society publicou as duas primeiras revistas em 1665, *Journal de Sçavans* e *Philosophical Transactions of the Royal Society of London* (MEADOWS, 1999).

contribuiu para o desenvolvimento da ciência e para a abertura de novos horizontes de pesquisa.

No Brasil, essa atividade está muito relacionada aos cursos de pós-graduação, conduzidos pelas universidades e institutos públicos de pesquisa, e se desenvolveu a partir da década de 1970, após a promulgação da Lei n. 5.540, em 1968, da Reforma do Ensino Superior (BRASIL, 1968), norteadas pelas ideias básicas estabelecidas ao fim da Segunda Guerra Mundial sob a inspiração de um dos países vencedores, os Estados Unidos (GUIMARÃES, 2002).

Foi, ainda, ao longo de toda a década de 1970 que foram criados cerca de 800 novos cursos de mestrado e doutorado. Indiscutivelmente, a criação dos cursos de pós-graduação contribuiu e impulsionou as atividades de pesquisa e a consequente produção de conhecimento, além de responder pela formação de recursos humanos altamente qualificados (GRACELLI e CASTRO, 1985). Estes autores, ao analisar o desenvolvimento da pós-graduação no Brasil, comentam que esta começa muito tarde, um século depois da norte-americana, porém historicamente em um momento propício no país, época de grande dinamismo na sociedade e na economia, motivo pelo qual teve um ritmo surpreendente de crescimento (GRACELLI e CASTRO, 1985). Alguns fatos importantes são relacionados por estes autores como ações motivadoras para o crescimento, como: a institucionalização de pós-graduação com a Lei n. 5.540/68 da Reforma do Ensino Superior (BRASIL, 1968), as medidas institucionais para aprimoramento do quadro de docentes, com a exigência de titulação para a promoção na carreira bem como os incentivos das agências de fomento para a criação desses espaços institucionais (GRACELLI e CASTRO, 1985).

SCHWARTZMAN (2010), no texto preparado como subsídio para o planejamento do Plano Nacional de Pós-Graduação (PNPG) relativo ao período de 2011-2020, também comenta que:

Com cerca de 10 mil doutores sendo formados a cada ano a Pós-Graduação brasileira atingiu um tamanho considerável, aproximando-se, em termos absolutos, de países de grande porte e altos níveis de desenvolvimento econômico. Além disso, o número de publicações científicas de autores brasileiros vem aumentando

continuamente. Esses resultados se explicam pela combinação de alguns fatores que incluem: a) a adoção do modelo norte-americano de cursos de pós-graduação estruturados a partir da reforma universitária de 1968; b) as exigências legais de titulação para a contratação e promoções nas carreiras universitárias, e para o reconhecimento e autorização de cursos e universidades; c) os investimentos das agências de fomento como a CAPES, CNPq, FAPESP outorgando bolsas de estudos no Brasil e no exterior; d) o sistema de avaliação de qualidade instituído pela CAPES (SCHWARTZMAN, 2010).

Isto é, quase todas as ações citadas em 1985, que motivavam o crescimento da pós-graduação, continuam agindo como mecanismos motivadores do bom desempenho da pós-graduação brasileira.

Esta boa atuação também é evidenciada em um estudo realizado, sobre a formação dos doutores nos cursos de Pós-graduação, no país (VIOTTI, 2010), no livro *Doutores 2010: estudos da demografia da base técnico-científica brasileira*.

O comentário sobre este estudo detalhado e pioneiro no Brasil neste trabalho se justifica pelo fato de haver uma forte ligação entre a titulação dos doutores e a produção científica, e conforme pode ser verificado, um dos objetivos deste trabalho é identificar a produtividade das teses em nível de produção de artigos do Programa de Pós-graduação do IPEN.

Conforme demonstrado pelos estudos, VIOTTI (2010), comenta:

Na verdade, o objetivo de dotar o Brasil de uma adequada população de doutores vem sendo perseguido com determinação há muitos anos. O crescimento de cerca de mil por cento, ocorrido no número de doutores titulados anualmente entre 1987 e 2008, evidencia esse fato. Essa é uma das razões pelas quais é possível classificar a política de implantação, expansão e controle de qualidade dos programas de mestrado e doutorado no Brasil, que se estruturou em meados dos anos 70, como um raro exemplo de política de estado. Essa política demonstrou uma continuidade surpreendente ao longo das últimas décadas, independentemente das inúmeras mudanças de governo e até mesmo de regimes políticos ocorridos no período (VIOTTI, 2010, p. 19).

VIOTTI (2010), também comenta que inferências sobre a intensidade e qualidade da produção científica dos programas de doutorado, medidas por intermédio de indicadores bibliométricos, como número de artigos publicados em revistas indexadas e número médio de vezes em que são citados, certamente colocariam os programas brasileiros em posição de desvantagem em relação aos programas de universidades americanas,

Entretanto, prossegue o Professor VIOTTI (2010), deve-se levar em conta a juventude dos programas brasileiros, e em comparação com outros países o Brasil está em posição confortável, com crescimento expressivo nos últimos anos. Este desempenho certamente deve-se à forte atuação da CAPES; desde a década de 1970 quando implantou o sistema de avaliação dos cursos de pós-graduação, este mecanismo tem-se mostrado forte indutor da produção científica de professores e alunos dos programas de doutorado no país.

A partir de 1995, a Pós-graduação passa a se expandir de forma vigorosa. A titulação de doutores em programas credenciados pela CAPES vai de aproximadamente 2 mil, naquele ano, para 11.500 em 2009, com taxa anual média acima de 12%. Se mantidas essas taxas médias anuais de crescimento, em 2020 se titulariam 40 mil doutores (PALIS JUNIOR, 2010).

Porém, VIOTTI (2010), alerta que apesar dos avanços, o estudo realizado e apresentado em *Doutores 2010* evidencia que o contingente de doutores no país ainda é pequeno se comparado com os Estados Unidos, por exemplo, em 2008 foram titulados no Brasil aproximadamente 11 mil doutores (FIG. 3), enquanto nos Estados Unidos foram 48 mil titulados no mesmo período.

PALIS JUNIOR (2010) realizou uma análise da produção científica e apresenta dados bem atualizados a esse respeito, comenta que em números de artigos científicos publicados em revistas indexadas o crescimento tem sido notável, acima da média mundial. Os dados apresentados na FIG. 4 indicam o crescimento do número de trabalhos ao longo dos anos e seu percentual em relação aos trabalhos científicos de todo o mundo

Figura 3 – Número de doutores titulados no Brasil 1987-2008

Fonte: VIOTTI, E. B., 2010, Cap. 1, p. 22.

Os indicadores foram elaborados a partir da base de dados internacional *Thomson Reuters. National Science Indicators. Database – Standard ESI, 2008*.

PALIS JUNIOR (2010) apresentou esses dados do desempenho da produção científica para subsidiar as suas propostas de ampliar e sustentar o avanço da ciência brasileira até 2020. No texto ele sugeriu redobrar os esforços para que de forma sustentável as atividades de Ciência e Tecnologia contribuam decisivamente para o avanço científico-econômico do país, bem como para manter a presença brasileira no cenário internacional baseada na competência científica.

PENA (2010), refletindo sobre o estado das ciências básicas no Brasil e objetivando identificar medidas que possam contribuir para o seu crescimento qualitativo, comenta que é notável o crescimento da produção científica brasileira em periódicos indexados no período de 1981-2007, em comparação com o resto do mundo, entretanto quando se examina o ranking de citações (ou quanto os artigos

	Brazil		
Years	Papers Brazil (A)	Papers World (B)	% Papers (A)/(B)
1981-1985	11.560	2.442.133	0,47
1982-1986	12.199	2.517.331	0,48
1983-1987	12.584	2.569.103	0,49
1984-1988	13.091	2.627.219	0,50
1985-1989	13.798	2.705.569	0,51
1986-1990	15.028	2.768.437	0,54
1987-1991	16.391	2.830.785	0,58
1988-1992	18.417	2.931.764	0,63
1989-1993	20.193	3.015.294	0,67
1990-1994	22.253	3.135.917	0,71
1991-1995	24.640	3.271.659	0,75
1992-1996	27.290	3.413.565	0,80
1993-1997	29.958	3.517.849	0,85
1994-1998	34.113	3.632.560	0,94
1995-1999	38.919	3.704.376	1,05
1996-2000	43.332	3.744.359	1,16
1997-2001	48.203	3.787.141	1,27
1998-2002	53.722	3.834.492	1,40
1999-2003	59.132	3.950.613	1,50
2000-2004	64.030	4.025.939	1,59
2001-2005	71.205	4.227.462	1,68
2002-2006	78.898	4.412.101	1,79
2003-2007	85.453	4.592.036	1,86
2004-2008	101.263	4.865.868	2,08

Figura 4 – Número de trabalhos e seu percentual em relação aos trabalhos científicos no mundo
 Fonte: Thomson Reuters. *National Science Indicators. Database – Standard ESI*. CD-ROM 2008
 Fonte: PALIS JUNIOR, J. (2010), p. 79.

são citados), o que indica o impacto da ciência produzida, verifica-se que o país ocupa um modesto 24º lugar, o que significa que há um descompasso entre a produção científica bruta e a qualidade da ciência feita no país, isto é, descompasso

entre a força científica brasileira e o grau de influência internacional. Cita cinco fatores sistêmicos que dificultam a realização da pesquisa de alta qualidade; entre eles a pesquisa baseada na pós-graduação, e mostra como exemplo os dados apresentados na FIG. 5, onde se observa que a produção científica brasileira tem crescido essencialmente na mesma proporção do desenvolvimento dos doutores titulados na Pós-graduação.

Figura 5 – Crescimento da titulação de doutores e número de artigos publicados
Gráfico preparado a partir de *National Science Indicators, USA Base Standard – ESI* (2007)
Cortesia de Guimarães, J. (2010).
Fonte: PENA, S. D. (2010).

Na visão de PENA (2010), o problema de basear a pesquisa no trabalho dos alunos de pós-graduação é que estes têm um prazo limitado fixado pelos cursos de pós-graduação e pelas agências de fomento para finalizar seus projetos, esta premência favorece o baixo grau de inovação dos artigos, embora haja exceções. PENA (2010) complementa os outros quatro fatores negativos: a ênfase cientométrica em números absolutos; ênfase em rápido retorno tecnológico da pesquisa; pulverização de recursos e a demanda criacionista das agências de fomento, e prossegue afirmando que embora estes fatores não possam ser eliminados porque fazem parte do ciclo virtuoso pelo qual a ciência brasileira tem progredido, é preciso rever algumas políticas que foram virtuosas no passado, porém que não se adequam mais.

3.1 A produção de indicadores em ciência

É bastante conhecida a afirmação de William Thompson, mais conhecido como Lorde Kelvin¹²: *a capacidade de mensurar e expressar algo em números permite um conhecimento mais profundo sobre determinado assunto*; sublinha, ao mesmo tempo, que *a incapacidade de expressar qualquer coisa em números é prova fidedigna de um conhecimento pobre e insatisfatório* (THOMPSON, 1889 citado por GEWANDSZNAJDER, 2010). O ato de medir está presente na vida do homem, desde o início dos tempos, é uma atividade rotineira na cultura humana.

GEISLER (2000); MEADOWS (1999), comentam que nas ciências exatas, mais especificamente na física e na engenharia, os processos dispõem de uma quantificação muito bem elaborada, aos aspectos relevantes dos processos naturais é possível associar grandezas capazes de serem medidas em termos de unidades com instrumentos apropriados.

Analisando as relações causa e efeito entre a produção científica e os fatores que a influenciam observa-se que estas são eminentemente probabilísticas (estocásticas). TRZESNIAK (1998), ressalta que as características das relações causa e efeito podem ser determinísticas ou estocásticas; em uma relação determinística a causa e efeito tem uma ligação direta, a presença de **a** implica necessariamente o surgimento de **b**; já na relação estocástica, a causa e efeito tem uma ligação indireta, fortemente associada à probabilidade ou à variação de o fenômeno acontecer.

O que são indicadores? O volume de definições para indicadores é expressivo e varia de um autor para outro. A organização governamental Financiadora de Estudos e Projetos (FINEP) é grande usuária e produtora de indicadores e considera que estes são: *especificação quantitativa e qualitativa para medir o atingimento de um objetivo* (FINANCIADORA DE ESTUDOS E PROJETOS, 2002).

¹² Lorde Kelvin (1824-1907), físico, matemático, teve uma contribuição fundamental para o campo da termodinâmica, criador da escala Kelvin.

Já a Instituição intergovernamental Organização para a Cooperação e Desenvolvimento Econômico (OCDE) estabelece que indicadores são uma série de dados definidos para responder a perguntas sobre um fenômeno ou um dado sistema (ORGANISATION DE COOPERATION ET DE DEVELOPEMENT ECONOMIQUES OECD, 2001)

Para a *International Standardization Organization (ISO)* os indicadores são expressão numérica, simbólica ou verbal empregada para caracterizar as atividades em termos quantitativos e qualitativos objetivando determinar seu valor (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION - ISO, 1998).

Depreende-se dessas definições que indicadores são unidades de medida que permitem aferir quantitativamente os resultados e impactos de processos e projetos. Estes instrumentos de medição pertencem ao contexto da gestão, uma vez que a medição em geral nunca é realizada em vão, sempre há uma finalidade maior, qual seja, a avaliação, os indicadores então são subsídios para o processo de avaliação. Em termos gerais, os indicadores têm ampla gama de aplicação e uso, são parâmetros utilizados em qualquer ramo de atividade. Para avaliar especificamente a atividade científica utilizam-se os indicadores bibliométricos, foco deste trabalho, baseados na análise estatística dos dados fornecidos pela atividade científica.

No entanto, a seleção e a construção de indicadores de produção científica (*output*) são uma tarefa complexa, cada área de conhecimento possui características próprias, principalmente em relação à divulgação dos seus resultados. Assim, a área de ciências exatas e biológicas não tem nem a mesma cultura nem o mesmo ritmo de produção, tampouco utiliza os mesmos canais de divulgação que as ciências sociais. Portanto, qualquer modelo de avaliação deve considerar tais diferenças (KOBASHI e SANTOS, 2004).

3.1.1 Histórico dos indicadores

Em nível mundial, para entender o desenvolvimentos dos indicadores, é preciso retroagir à época dos anos 1950. Após a Segunda Guerra Mundial,

constatando a relevante atuação da ciência e da tecnologia, BUSH (1945), apresenta para o presidente dos Estados Unidos da época o relatório *Science: the endless frontier*, o qual teve grande repercussão e norteou o estabelecimento de uma política científica e tecnológica nos Estados Unidos, e que logo se estenderia para outros países. Esse relatório sintetizava o otimismo e a confiança no papel e desenvolvimento da ciência e tecnologia, e defendia a ideia de que a ciência deveria ser uma prática autônoma, autorregulável e subsidiada pelo governo, neste modelo a política científica e tecnológica só tinha que assegurar a entrada de recursos financeiros (*input*), garantindo assim a justificativa social da ciência e o investimento público nela. A este modelo se denominou *Modelo Linear* (CANCHUMANI, 2010).

Porém, diante da crescente demanda por recursos e espaço para o desenvolvimento das atividades de Ciência e Tecnologia, os Estados Unidos propõem um relativo controle e avaliação da produção de ciência objetivando garantir sua qualidade. Foi esse cenário também que despertou a atenção de analistas e estudiosos e que levou a *Organization for Economic Co-operation and Development* (OECD), em 1957, a reunir um grupo de especialistas com o objetivo de estudar gastos e apresentar propostas para a mensuração da atividade científica bem como buscar a padronização, tanto na coleta como na emissão dos dados estatísticos, de forma que permitisse uma comparação com outros países (ORGANISATION DE COOPERATION ET DE DEVELOPEMENT ECONOMIQUES, 2001).

Em 1963, o grupo de especialistas se reuniu na Vila Falconieri de Frascati, na Itália, para discutir e revisar o documento *Proposta de um sistema padrão para avaliação em pesquisa e desenvolvimento*. Assim nasce a primeira versão oficial do manual Frascati¹³. O manual avalia a eficácia da produção de ciência e tecnologia mensurada por meio dos investimentos e gastos realizados em pesquisa e desenvolvimento de ciência e no número de patentes para a produção tecnológica. Apresenta as principais classificações setoriais e funcionais dos dados de P&D e discorre sobre a medição e classificação principalmente no que se refere às

¹³ Manual Frascati: <http://www.mct.gov.br/upd_blob/0023/23423.pdf>.

estimativas e aos ajustes necessários; dedica também um capítulo à comparação desses dados com outras variáveis econômicas e à comparação internacional do dispêndio em Pesquisa e Desenvolvimento. (ORGANISATION DE COOPERATION ET DE DEVELOPEMENT ECONOMIQUES, 2001).

Já na década de 1990 e na busca por otimizar o ajuste entre ciência e tecnologia, empresa e mercado, o investimento público começa a ser canalizado para a criação de centros de transferência de tecnologias em universidades, empresas, visando otimizar o processo de inovação. Assim, em 1992 é apresentada pela OCDE a primeira edição do *Manual de Oslo*, objetivando que ele sirva de guia para a compilação de dados relativos à inovação tecnológica. Esses indicadores, basicamente, são informações de empresários coletadas por questionários para mensurar o nível de aproveitamento no mercado das descobertas científicas (CANCHUMANI, 2010).

A OCDE organizou e publicou ainda outros manuais, cada um com um foco especial objetivando sempre o desenvolvimento de conceitos, metodologias e técnicas para a construção de indicadores, e que ficaram conhecidos como *Manuais da Família Frascati*, como por exemplo o *Manual da Canberra*, cujo objetivo é difundir a padronização de uma estrutura conceitual comum para a compilação, a análise de perfis e as tendências de dados relativos aos estoques e fluxos de recursos humanos em ciência e tecnologia. Para ilustrar a relação de manuais, apresenta-se a FIG. 6.

Tipo de dados	Título
A. A “Família Frascati”	Série: <i>Medição das atividades científicas e tecnológicas</i>
ID	Manual de Frascati: proposta de normas práticas para inquéritos de investigação e desenvolvimento experimental
	<i>Estatísticas de ID e medidas de output no sector do ensino superior. “Suplemento do Manual de Frascati”</i> (OCDE, 1989b)
Balança de pagamentos tecnológica	<i>“Manual para a medida e interpretação da balança de pagamentos tecnológica – Manual BPT”</i> (OCDE, 1990)
Inovação	<i>Diretivas propostas para a recolha e a interpretação de dados sobre inovação tecnológica – Manual de Oslo</i> (1997a)
Patentes	“Utilização dos dados de patentes como indicadores de Ciência e Tecnologia – Manual de Patentes” (OCDE, OCDE/GD (94)114, 1994b)
Pessoal de C&T	“Manual sobre a medida dos recursos humanos dedicados à ciência e à tecnologia – Manual de Canberra” (OCDE, 1995)
B. Outras obras metodológicas para medir a ciência e a tecnologia	
Alta tecnologia	Revisão das classificações dos setores e dos produtos de alta tecnologia (OCDE, Documentos de trabalho da STI 1997/2)
Bibliometria	“Recomendações para a utilização de indicadores bibliométricos e análise dos sistemas de investigação: métodos e exemplos”, por Yoshiko Okubo (OCDE, Documentos de trabalho da STI 1997/1)
Globalização	<i>Manual de indicadores de globalização económica</i> (título provisório, em preparação)
C. Outras obras estatísticas aplicáveis da OCDE	
Estatísticas de ensino	<i>Manual de estatísticas comparativas de educação</i> (em preparação)
Classificação do ensino	<i>Classificação dos sistemas de educação. Manual de utilização da ISCED-97 nos países da OCDE</i> (OCDE, 1999)
Estatísticas de formação	<i>Manual do melhor método para a recolha de estatísticas de formação – Conceitos, medida e inquéritos</i> (OCDE, 1997b)

Figura 6 – Manuais metodológicos da Organization for Economic Co-operation and Development (OCDE)
 Fonte: Reproduzido de: *Organization for Economic Co-operation and Development Manual Frascati*, 2002, p. 24.

O Manual de Patentes trabalha com patentes de invenção industrial, não abordando propriedades intelectuais, como marcas, desenhos e o direito autoral. O

manual estabelece diretrizes sobre como os dados de patentes podem ser utilizados em análises que os relacionam com outras estatísticas sobre atividades científicas, tecnológicas e econômicas CANCHUMANI (2010); LIBERAL (2005).

No Brasil, o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), do Ministério da Ciência e Tecnologia (MCT), foi a instituição que primeiro realizou esforços para gerar indicadores de Ciência e Tecnologia (C&T) para o País. A partir dos anos 1980, o CNPq iniciou a coleta e a publicação de informações sobre os recursos do governo federal aplicados em C&T, seguindo as primeiras recomendações do Manual Frascati da OCDE, para os gastos em Pesquisa e Desenvolvimento (P&D), e as sugestões da Organização das Nações Unidas para a Educação, Ciência e Cultura (UNESCO), para as atividades científicas e técnicas correlatas (ACTC). Uma década depois, os mesmos procedimentos passaram a ser utilizados na maioria dos estados brasileiros, permitindo a obtenção de um quadro abrangente dos recursos públicos aplicados em C&T (BRASIL, 2002).

A partir de 1999, o MCT assumiu a responsabilidade pela organização e divulgação das informações de C&T do País, de forma centralizada. Inicialmente, o levantamento e a divulgação dos indicadores concentravam-se no que passou a se denominar indicadores de insumo (*input*), isto é, no dimensionamento dos recursos financeiros e humanos investidos em ciência e tecnologia, ou seja, a mensuração se limitava à identificação dos recursos aplicados à pesquisa. Mais recentemente, foram desenvolvidos os chamados indicadores de resultados, inicialmente limitados à produção científica e, posteriormente, incorporando a produção de patentes e a transferência de tecnologia entre países.

Estas iniciativas são recentes principalmente no que se relaciona à construção de indicadores de impacto sabidamente de construção mais complexa, isto é, como mensurar se determinado resultado científico ou tecnológico afeta as várias dimensões das condições de existência dos indivíduos, seja no próprio campo científico e tecnológico, seja na dimensão econômica ou social.

Os indicadores de impacto na dimensão científica e tecnológica são atualmente os mais desenvolvidos, em especial aqueles construídos no campo da bibliometria (BRASIL, 2002).

Para os países menos desenvolvidos o desafio não é apenas ampliar a cobertura e o escopo dos indicadores de C&T, preservando a comparabilidade internacional, mas também melhorar a qualidade e a representatividade das informações primárias por meio de investimentos na coleta e no tratamento dos dados. O desenvolvimento de estudos nacionais permite conhecer melhor as características próprias das estruturas de C&T e, com isso, refinar as metodologias utilizadas na produção de indicadores (BRASIL, 2002).

Na América Latina, as instituições que participam da Rede Ibero-americana de Indicadores de Ciência e Tecnologia (RICYT), Organização dos Estados Americanos (OEA) entre outras instituições, após a realização de três seminários internacionais, apresentaram à comunidade científica o Manual para a Normalização de Indicadores de Inovação Tecnológica para a América Latina e o Caribe, denominado *Manual de Bogotá*, o qual é um instrumento de orientação teórica e metodológica que procura responder à necessidade de compreender as especificidades dos processos de inovação na América Latina e no Caribe. Inspirado no *Manual de Oslo* revela a preocupação dos seus produtores com os critérios e procedimentos que assegurem a comparabilidade regional e internacional dos indicadores aí considerados (JARAMILLO, 2001).

Para analisar a construção de indicadores de ciência e tecnologia em determinado tempo e espaço, é importante ter presente a complexidade que envolve os processos científicos e tecnológicos, qualquer decisão racional em matéria de avaliação envolve a análise de inúmeras variáveis e os especialistas nesta área, como VELHO (1999), recomendam cautela na análise e na emissão de resultados.

GUIMARÃES (2002), comenta que a avaliação da atividade científica tem sido realizada no país principalmente para permitir a distribuição de recursos financeiros. Foram criadas algumas metodologias cujo conceito-chave é a qualidade e seu instrumento de legalidade é a quantificação. A avaliação nesses moldes é usada

como instrumento de tomada de decisão e justificativa racional e objetiva na administração de recursos destinados à pesquisa.

Para MACIAS-CHAPULA (1998), os indicadores da atividade científica estão no centro do debate nas relações do avanço da ciência e da tecnologia e no progresso econômico e social do país. Não se admitem revisões ou reformulações de políticas científicas sem se recorrer aos resultados dos indicadores existentes.

MORAVCSIK (1988) e SANCHO (1990), sublinham que o processo científico se assemelha aos processos que se desenvolvem nas empresas com insumos (*input*) – processos – produtos (*output*), em que custos e investimentos são itens tangíveis. Portanto, os indicadores de ciência devem ser divididos nessas duas categorias. Todavia, benefícios e resultados da ciência são intangíveis; não são passíveis de se contabilizarem em termos econômicos ou com modelos convencionais, o que demanda o desenvolvimento da utilização de medidas indiretas com métodos apropriados.

Até meados de 1970, o foco das avaliações restringia-se quase exclusivamente à coleta de estatísticas relacionadas com os insumos (*input*), como verbas atribuídas ao número de pessoas envolvidas em Pesquisa e Desenvolvimento (P&D); isso se deveu, segundo VELHO (1999), ao fato de se acreditar que tais informações fossem suficientes para se estabelecer uma política científica. Atualmente, o interesse está voltado para os indicadores de resultados (*output*), publicações, produtos, assim a elaboração e o uso de indicadores bibliométricos e cientométricos tornaram-se essenciais. SPINAK (1998), comenta que a mensuração dos resultados (*output*) se constitui na segunda metade da tarefa e que apresenta um maior nível de complexidade.

Com relação ao uso dos termos *input* e *output*, faz-se necessário um comentário. Esses termos se referem aos insumos e aos produtos resultantes dentro do processo de construção de ciência, porém são mais comumente adotados no idioma inglês na maioria dos estudos que envolvem mensuração. A ideia de entrada e saída de resultados é adotada por autores como SPINAK (1998); SANCHO (1990),

e VELHO (1999), os quais fazem uma analogia da atividade científica com os processos produtivos que se desenvolvem em empresas.

Para POBLACIÓN e OLIVEIRA (2006), na geração de conhecimento científico, de forma bem generalizada podem-se considerar esses dois grandes momentos, sendo que o *input* é uma combinação dos fatores que viabilizam a produção de determinada quantidade de bens e serviços, o *output*. Assim a mensuração destes dois aspectos do processo é a base para a elaboração dos indicadores científicos.

Assim, exemplificando alguns itens, são denominados *inputs*, dentro do processo de construção de ciência: os recursos humanos, a infraestrutura, o financiamento e a política científica; entre os *outputs* da atividade científica, um dos mais relevantes é a produção científica (artigos, teses, livros, patentes), a produção tecnológica, a transferência de tecnologia entre outros (POBLACIÓN et al, 2009). A importância dos insumos (*inputs*) no processo de geração de novos conhecimentos está diretamente relacionada aos resultados (*output*), podendo influenciar na qualidade e quantidade desses dois aspectos (NORONHA e MARICATO, 2008).

Pelos comentários dos autores citados conclui-se que, a produção científica vista por essa abordagem de insumos/produtos ou resultados demanda instrumentos de medida que gerem indicadores científicos, como meio de avaliar os resultados dos investimentos realizados em ciência no âmbito das instituições, bem como para que os indicadores se tornem ferramentas úteis devem ser produzidos com regularidade, permitindo a formação de séries temporais que possibilitem visualizar as tendências no tempo, os dados coletados devem permitir comparações, como também devem ser passíveis de enfocar qualquer área de conhecimento.

GEISLER (2000), afirma que a mensuração por meio de indicadores é o método mais apropriado para avaliar ciência e tecnologia, porém é necessário estabelecer múltiplos indicadores para conseguir resultados mais adequados aos diversos e complexos aspectos da ciência e tecnologia.

VINKLER (1988), define os possíveis tipos e níveis de indicadores dependendo do objetivo e da abrangência em: *micro*, *meso* e *macro*: no nível *micro*,

por exemplo, os indicadores sobre publicações podem se referir a uma só pessoa; no nível *meso*, a um grupo de trabalhos publicados no âmbito de uma Instituição; no nível *macro*, à análise de todos os trabalhos publicados de determinado campo científico (FIG. 7).

Tipos de indicadores	Nível de indicadores		
	Micro	Meso	Macro
	Pessoas	Institutos	Grupos de institutos
Organização	Grupos	Departamentos	Países
			Mundo
Temáticos	Projetos	Campos e subcampos temáticos de áreas de pesquisa	Disciplinas científicas ou as ciências como um todo
Publicação	Um trabalho (<i>paper</i>) científico publicado	Grupo de trabalhos científicos publicados (<i>papers</i>)	Todos os trabalhos publicados

Figura 7 – Tipos de nível de indicadores
 Fonte: VINKLER, P. *Bibliometric indicators for scientometrics*, 1988.

O emprego de indicadores bibliométricos apresenta uma série de vantagens para sua utilização na avaliação científica, por se tratar de dados verificáveis, reproduzíveis, que apresentam resultados objetivos, numéricos, além de poderem ser aplicados a um grande volume de dados.

As características dos indicadores têm sido estudadas por vários autores, entre eles SANZ CASADO e MARTIN MORENO (1997) destacam em particular, três características: a parcialidade, a convergência e a relatividade.

A primeira se refere ao fato de que cada indicador mostra um aspecto de avaliação que está sendo realizado. A segunda diz respeito a que todos os indicadores convergem para proporcionar um amplo conhecimento da atividade em questão, razão pela qual os pesquisadores recomendam utilizar um número maior de indicadores para não se correr o risco de se obterem resultados tendenciosos; uma vez que a atividade científica é multidimensional, não pode ser analisada por um só indicador.

Por último, a informação que se obtém por meio dos indicadores é relativa à disciplina estudada ou abordada, não deve ser extrapolada para outra disciplina, já

que os hábitos de pesquisa dos grupos são distintos e sua replicação demanda uma análise criteriosa de aplicabilidade e conveniência.

Com relação à classificação dos indicadores, SANZ CASADO e MARTIN MORENO (1997), com base em técnicas de estatística, propõem a seguinte denominação:

- ✓ indicadores unidimensionais: referindo-se àqueles que utilizam técnicas de estatística univariável; tratam de um só aspecto ou uma característica das publicações científicas ou dos pesquisadores dependendo do objeto de análise, sem identificarem ou relacionarem vínculos que possam existir entre eles;
- ✓ indicadores multidimensionais (ou bidimensionais ou relacionais): que utilizam estatística multivariável; estudam simultaneamente várias características dos documentos, estabelecem correlação entre os documentos ou mostram hábitos dos pesquisadores; essas correlações ou estruturas são mostradas graficamente por meio dos mapas da ciência, ou mapas bibliométricos.

Estas características também são registradas por BRAAM et al. (1991); VAN RAAN (1993, 2001, 2005); NOYONS (2001).

SPINAK (1998); MACIAS-CHAPULA (1998) e COURTIAL (1994), dividem ainda os indicadores em:

- ✓ indicadores de produção científica, obtidos pela contagem do número de publicações por tipo de documentos (ou nível monográfico): livros, artigos, relatórios, capítulos de livros; ou ainda por instituição, por áreas de conhecimento; são indicadores quantitativos;
- ✓ indicadores de citação, muito utilizados e conhecidos, são construídos por meio da contagem do número de citações que um artigo, autor, periódico instituição ou país recebe. Considerados o meio mais reconhecido de se atribuir crédito a um autor; são indicadores de qualidade;

- ✓ indicadores de ligação são os que utilizam, fortemente, técnicas de estatística mais avançada de agrupamento; são construídos por meio do levantamento de itens de ligação ou coocorrência, como palavras, autores, áreas temáticas, índices classificatórios. São utilizados para elaborar redes de relacionamento, clusters, mapas de estruturas do conhecimento. São indicadores que mesclam a abordagem quantitativa e qualitativa, apresentado resultados interessantes.

3.1.2 Os mapas da ciência

Com relação aos mapas da ciência, estes se situam no âmbito das diversas técnicas que a bibliometria avançada utiliza, e surgiram há mais de três décadas. Durante algum tempo essa técnica não teve muita repercussão, porém, graças ao desenvolvimento da tecnologia da informação, e ao aparecimento da Web em 1993, os *mapas da ciência* adquiriram uma interface gráfica mais amigável e sua visualização tornou-se mais eficiente, permitindo uma navegação dentro de determinada área de pesquisa, similar a um mapa geográfico que propiciou maior visibilidade, aceitação e desenvolvimento (NOYONS (1999, 2001); BUTTER et al., (2006).

Mapas da ciência são representações simbólicas de domínios científicos, ou organizações, nos quais os elementos são distribuídos pela similaridade de tal forma que os mais relacionados se situam mais próximos e os menos se localizam mais distantes (NOYONS, 2001). A produção científica pelo tema é um dos aspectos em que os mapas da ciência têm mais aplicação. Essa modalidade de estudo permite descobrir a evolução, mudanças, comportamento ou extinção das linhas de pesquisa ao longo do tempo, bem como tendências (NOYONS, 2001).

A construção dos mapas da ciência pode se basear na análise de coocorrências de itens específicos como: citações e palavras-chave ou simplesmente palavras dos textos, autores, publicações, áreas temáticas, e assim por diante. Estes elementos identificados em um volume expressivo de literatura científica de

determinada área de conhecimento possibilitam a identificação de estruturas abstratas, de uma forma visual mais rápida. Como nessa modalidade são utilizados diferentes elementos (campos) bibliográficos das publicações científicas para análise, cada um deles revela uma estrutura específica, dependendo da finalidade para a qual se está construindo o mapa (NOYONS, 1999).

Assim, por exemplo, um mapa baseado na coocorrência de autores – informação extraída do campo autor – vai revelar a estrutura social da área e o tipo de relacionamento que acontece entre autores daquela área estudada.

O grande número de campos utilizados – autor, título, filiação dos autores, classificação, indexação, resumo e outros – permite uma variedade significativa de estudos, por meio do uso de parâmetros bibliométricos para produzir informação confiável. No processamento e nas análises dos dados, utilizam-se softwares especializados e operações matemáticas que basicamente aplicam algoritmos, criam listas de frequência de termos (conceitos) e efetuam filtros retirando palavras supérfluas; utilizam-se também técnicas de visualização bidimensionais para apresentação dos mapas (NOYONS e VAN RAAN, 2002).

O mais interessante dessa técnica deve-se ao fato de ela conseguir revelar ou evidenciar dados que já se encontram contidos nos documentos, ou seja, não são passíveis de um pré-julgamento ou padrões estabelecidos *a priori* (NOYONS e VAN RAAN, 2002).

3.1.3 A técnica de coocorrência de palavras (*co-word analysis*)

Dentro das técnicas para construção de mapas da ciência, está a técnica de coocorrência de palavras, mais conhecida como *co-word analysis*.

Trata-se de uma técnica de análise de conteúdo, baseada na premissa de que a coocorrência de palavras estabelece a possibilidade de identificar o grau de “parentesco” entre os itens presentes em um texto analisado; assim, quanto mais **pares** de itens coincidentes houver, maior será a probabilidade de haver correlação entre esses itens (VAN RAAN, 1993).

Essa técnica foi desenvolvida inicialmente na década de 1980, pelo *Centre de Sociologie de l'Innovation da Ecole Nationale Supérieure* de Mines de Paris, em colaboração com o *Centre National de La Recherche Scientifique* (CNRS). O sistema foi chamado de LEXIMAPPE, e a técnica já foi empregada para mapear a dinâmica de desenvolvimento de várias áreas de pesquisa. Posteriormente, foi testada e utilizada por pesquisadores de outros países (HE, 1999).

Como base conceitual para dar sustentação a esse procedimento, neste trabalho utilizaram-se os estudos sobre a produção de indicadores multidimensionais de COURTIAL (1994); VAN RAAN (1993); NOYONS (1999); HE (1999); BASSELAAR e HEIMERIKS (2006); BASSECOULARD et al. (2007); YOON et al. (2010).

A maior parte dos trabalhos que utilizam essa técnica desenvolve representações gráficas onde se mostra o grau de centralidade e densidade das áreas estudadas por meio de redes de ligação. As palavras podem ser selecionadas dos textos, dos títulos, dos resumos, ou ainda ser utilizadas as palavras-chave atribuídas pelos indexadores (HE, 1999).

3.2. A bibliometria e a cientometria e a informetria

Em décadas passadas, a definição e a abrangência desses termos já estiveram mais discordantes, atualmente, devido ao grande número de trabalhos, ao envolvimento dos especialistas e à abrangência do objeto de estudo dessas áreas, eles se apresentam com uma delimitação mais clara, embora ainda se empreguem nos trabalhos científicos os termos bibliometria e cientometria indistintamente, em outras palavras, a definição conceitual dos termos ainda não está consolidada.

Para KOBASHI e SANTOS (2008), tanto um termo como outro consistem em analisar, com a ajuda de métodos estatísticos e matemáticos, os elementos de um *corpus* documental a fim de determinar as relações existentes entre esses elementos. Um estudo bibliométrico pode ter por objetivos:

- a) estabelecer a estrutura da área estudada;

- b) estabelecer o posicionamento relativo de um assunto, de um periódico, de uma Instituição;
- c) analisar a evolução cronológica das relações entre elementos e detectar os acontecimentos emergentes ou atípicos.

Esse estudo visa conhecer como se comporta uma propriedade em relação a outra já conhecida, dessa forma, no caso das publicações científicas, podem-se conhecer todos os autores que trabalharam em determinado assunto, os periódicos que publicaram sobre esse assunto, os autores considerados referência no assunto, e assim por diante. Relacionando-se essas propriedades, obtêm-se ou evidenciam-se informações que ainda não haviam sido percebidas (KOBASHI e SANTOS, 2008).

Por ordem de aparecimento, primeiro surgiu a bibliometria, no início do século XX. Composta de duas raízes, *biblio* (livro) e *metria* que deriva do grego *metrikos* (medida), desenvolveu-se inicialmente pela elaboração de leis empíricas sobre o comportamento da literatura. Originalmente ficou conhecida como *bibliografia estatística*, expressão cunhada por Hulme em 1923, porém foi a partir do artigo *Statistical bibliography or bibliometrics?*, de Alan Pritchard, publicado no volume 25 do *Journal of Documentation* em 1969, que o termo se consolidou e popularizou (PRITCHARD, 1969).

A *linha do tempo* da história da bibliometria se iniciou antes do século XX. Há relatos de estudos esporádicos e alguns autores atribuem a autoria do termo a Paul Otlet¹⁴, em 1934, para quantificar a ciência e os cientistas, em sua publicação *Traité de la Documentation*, no qual bibliometria seria “a parte definida da bibliologia que se ocupa da medida ou quantidade aplicada aos livros”. Certamente trata-se de uma primeira definição a qual PRITCHARD (1969) amplia como sendo: todos os estudos que visam quantificar os processos de comunicação escrita (PINHEIRO, 2008).

¹⁴ Paul Otlet (1868-1944), fundador da ciência da bibliografia e documentação, produziu o *Traité de Documentation* (1934) onde apresenta as bases para a organização da informação registrada. Criou em 1895 o Instituto Internacional de Bibliografia (IIB) com o objetivo de organizar as diferentes fontes de investigação científica e fornecer informação para a recuperação em qualquer documento publicado mundialmente. Criou, também, em 1899, o Sistema de Classificação Decimal Universal (CDU).

O termo *informetria* foi introduzido por Otto Nacke em 1979 porém se popularizou na conferência internacional de “Informetrics” em 1987, quando foi adotado pela Federação Internacional de Documentação (FID), que assim a definiu: o conjunto de atividades métricas relativas à informação, cobrindo a bibliometria e a cientometria (EGGHE e ROUSSEAU, 1990).

A informetria se ocupa de estudos relacionados a mensurações na Web, na Internet, das redes sociais (citação, coautoria); de uma forma geral pode-se afirmar que a era da *Sociedade da Informação* é responsável pelo crescimento da informetria; diz-se também que todos os termos, bibliometria, cientometria e mais recentemente webometria, seriam englobados pela informetria, pois este seria um termo mais amplo (EGGHE e ROUSSEAU, 1990).

Para PINHEIRO (2008), o termo mais abrangente seria cientometria, correspondendo a estudos métricos de natureza política econômica e social de C&T, enquanto a informetria, composta por métodos estatísticos e matemáticos, mediria a informação em qualquer suporte, abrigando a bibliometria.

Já a Webometria seria empregada para medir a informação utilizando mecanismos de busca e softwares específicos.

TAGUE SUTCLIFFE (1992) e VAN RAAN (2005), apresentam as seguintes definições para:

- ✓ bibliometria é o estudo dos aspectos quantitativos da produção, da disseminação e do uso da informação registrada; desenvolve padrões e modelos matemáticos para medir esses processos, usando seus resultados para elaborar previsões, alocação de recursos e apoiar a tomada de decisão;
- ✓ cientometria é o estudo dos aspectos quantitativos da ciência e tecnologia, como uma disciplina ou atividade econômica. A cientometria é um segmento da sociologia da ciência, aplicada no desenvolvimento e no planejamento de políticas científicas. Envolve estudos quantitativos das atividades científicas incluindo a publicação, sobrepondo-se, portanto, à bibliometria. O objetivo da cientometria é identificar domínios de interesse em que os assuntos estão

concentrados, e compreender como e quanto os cientistas se comunicam. Tem como objetos de estudo: disciplinas, áreas e campos científicos e tecnológicos, atividades científicas, cientistas, colégios invisíveis, dissertações e teses, e documentos tecnológicos (patentes).

Para TAGUE-SUTCLIFFE (1992), a informetria vai além, estudando todos os processos quantitativos da informação em geral, incorporando, utilizando e ampliando as fronteiras da bibliometria e da cientometria.

Entretanto, conforme comentado anteriormente, a conceituação dos termos não está consolidada e ainda para alguns autores como GLANZEL (2006), tanto um como outro podem ser empregados indistintamente.

Embora o desenvolvimento e a aplicação das técnicas bibliométricas datem de décadas anteriores, sua aplicação acontecia principalmente nas atividades de organização, controle e serviços de bibliotecas, como a aplicação da *lei de Bradford*, sobre a frequência de distribuição de artigos em revistas, estabelecendo uma relação útil também para o processo de aquisição de material bibliográfico que ficou conhecida como a *lei de dispersão de Bradford*; o enunciado desta lei mostra que se os periódicos forem dispostos em ordem decrescente de produtividade de artigos de determinado assunto poder-se-á determinar o núcleo de títulos mais pertinentes ao assunto estudado; esta lei é útil para estudos de uso de coleção, em bibliotecas, uma vez que auxilia a identificar o núcleo básico de periódicos para aquisição, circulação e acervo, descartes e outras decisões administrativas (NORONHA e MARICATO, 2008).

O estudo de Lotka sobre produtividade de autores enuncia que a relação entre o número de autores e o número de artigos publicados por esses, em qualquer área científica, segue a lei do inverso do quadrado $1/n^2$. Isto quer dizer que analisando um dado período de tempo, um número n de artigos, o número de cientistas que escrevem dois artigos seria igual a $\frac{1}{4}$ do número de cientistas que escreveram dois, aqueles que escreveram três artigos seria igual a $\frac{1}{9}$ do número de cientistas que escreveram um, e assim sucessivamente (VOOS, 1974). Lotka demonstrou com esta lei que uma grande quantidade de literatura científica é

produzida por um pequeno número de autores e um grande número de pequenos produtores se iguala, em produção ao reduzido número de grandes produtores, ou simplificando: *muitos produzem pouco e poucos produzem muito*.

A terceira lei de Zipf descreve a relação entre a ocorrência de palavras nos textos científicos, a partir da qual são determinadas as palavras mais usadas que indicam o assunto do documento. Este enunciado está diretamente relacionado com a indexação temática automática de documentos científicos (NORONHA e MARICATO, 2008).

Porém, ao se historiar a bibliometria, a cientometria ou qualquer métrica para a ciência, deve-se mencionar Derek de Solla Price, professor inglês, Ph.D. em Física e em História, que em 1961 publicou os clássicos *Science since Babilon* e dois anos depois *Little science, big science*, que de acordo com BRAGA (1974), são obras de leitura “compulsória” para todo cientista.

Nestas obras ele trata da filosofia da ciência, da necessidade da pesquisa empírica sobre a atividade científica, dos processos de comunicação e da produtividade científica por meio de leis internacionalmente aceitas como: frentes de pesquisa, colégios invisíveis, crescimento exponencial da ciência, elitismo, e ainda da utilização sistemática de métodos quantitativos em especial de indicadores e modelos matemáticos. Essas obras definem e constituem marcos históricos para o que se denominou “ciência da ciência”, isto é, ciência para estudar a ciência.

Seu trabalho como pesquisador, cientista e tecnólogo está concentrado em mais de 300 documentos publicados ao longo de mais de 35 anos de pesquisas (BRAGA, 1974).

Retomando a reflexão sobre a importância dos métodos bibliométricos, RIP (1988) escreveu que o entendimento por parte da sociedade e dos governos da importância da pesquisa científica bem como os altos orçamentos demandados para essas atividades científicas despertavam o interesse em formas de estabelecer uma avaliação sistemática da ciência. Habitualmente a avaliação era realizada pelos próprios pesquisadores no sistema de *peer review*, mas esta avaliação era qualitativa e dizia respeito aos aspectos do estado da arte cognitiva de determinados campos

de pesquisa científica. Os administradores de pesquisa científica necessitavam de outra modalidade de informação, aquela em nível organizacional, dados contáveis ou informações quantitativas, em que os indicadores se apresentam como ferramentas complementares e valiosas para dar suporte a este outro segmento. Neste aspecto os indicadores científicos produzidos pelos métodos bibliométricos são os que tiveram maior desenvolvimento e um grande avanço para subsidiar as políticas científicas, embora os métodos bibliométricos para a gestão de informação continuem a serem utilizados.

Em síntese, se observa que a abrangência de estudo das disciplinas acima citadas é bastante próxima, estas podem ser consideradas complementares e têm o mesmo objeto de estudo, isto é, os aspectos quantitativos da ciência. Como a ciência em si é imensurável, medem-se os seus produtos, a citação desses produtos, os meios pelos quais ela se dissemina, são mensurações indiretas na tentativa de compreender a ciência como um corpo de ideias e fenômenos (VAN RAAN,1993).

No Brasil, podem-se identificar duas principais fontes de estudo na história do desenvolvimento dos estudos bibliométricos, uma é a partir do Instituto Brasileiro de Informação Científica e Tecnológica (IBICT) e a outra a partir de pesquisadores individuais de outras instituições interessadas em desenvolver os estudos em metrias da ciência, com forte tendência de formação de núcleos de pesquisas (PINHEIRO, 2008).

Os estudos métricos desenvolvidos inicialmente no IBICT contemplavam a aplicação das técnicas bibliométricas visando à gestão da informação e isto se justifica porque o IBICT atua em duas frentes, tanto como Instituto de Pesquisas voltado para a informação científica e tecnológica e a ciência da informação como com a tarefa de prover infraestrutura, serviços e produtos de informação (PINHEIRO, 2008).

O responsável pela introdução dos estudos utilizando as técnicas bibliométricas foi Tefko Saracevic da *Case Western Research University* nos Estados Unidos, o qual atuou como professor no primeiro programa de Mestrado em Ciência

da Informação no antigo Instituto Brasileiro de Bibliografia e Documentação (IBBD), que em 1976 passou a se denominar Instituto Brasileiro de Informação Científica e Tecnológica (IBICT). A disciplina ministrada pelo professor Saracevic era Processamento de dados na Documentação e incluía a bibliometria. Na época era hábito convidar professores estrangeiros para ministrar as disciplinas, uma vez que ainda não havia muitos brasileiros para atuar como orientadores.

A partir das orientações desse grupo de professores foi se formando uma massa crítica de pesquisadores no país que deu continuidade a trabalhos sobre este tema, publicando artigos e apresentações em eventos. PINHEIRO (2008), relata que o foco dos estudos métricos desenvolvidos no IBICT inicialmente estava voltado para a gestão da informação, enquanto o segundo grupo focava os estudos nos aspectos da política em Ciência e Tecnologia (C&T) pela abordagem da cientometria e a da geração de indicadores. Posteriormente, esta divisão caminha para a convergência, porque os profissionais do IBICT também direcionaram seus estudos para os aspectos da política em C&T.

O segundo grupo atuante nos estudos métricos no país é composto por professores e pesquisadores, atuantes nas diversas instituições brasileiras. É formado por dois conjuntos de profissionais, aqueles que não atuam em Ciência da Informação, mas desenvolvem pesquisas em metrias da comunicação científica e indicadores de C&T, e o outro conjunto de pesquisadores que atua na área de ciência da informação com as mais diversas graduações. Os pesquisadores destes dois conjuntos destacam-se pela atuação relevante na pesquisa, orientação de alunos e produção científica sistemática no tema o que permite que o país marque presença internacional, nesta área de estudo (PINHEIRO, 2008).

O que pode ser observado por meio do relato de PINHEIRO (2008) é que há no país uma convergência e fortalecimento dos estudos neste campo de estudo - metrias da comunicação científica -, que se comprova pelo crescimento dos núcleos de pesquisa formados no país, pelas publicações no tema, pelos eventos realizados. É visível também a diversidade de graduação dos pesquisadores atuantes, o que

enriquece a abordagem dos temas e contribui para o avanço e aprofundamento deste campo de estudo.

3.3 A literatura cinzenta (LC) e as Dissertações e Teses

No universo da documentação, as dissertações e teses são consideradas itens de *literatura cinzenta* (LC). Esclarece-se que a cor cinza não tem nenhuma conotação negativa, ao contrário, trata-se de um tipo de literatura que veicula, na maioria das vezes, informação de qualidade em primeira mão; apresenta originalidade e relata o progresso e o avanço de projetos científicos em determinadas áreas, além de ser considerada literatura diferenciada nessas áreas. Quanto à origem da denominação, esta parece estar relacionada ao antigo costume inglês de atribuir cores e termos para diferenciar os documentos (CARVALHO, 2001).

A literatura cinzenta também denominada literatura ***não convencional***, muito antes de receber essa denominação, que somente aconteceu em 1977, já era produzida por todos aqueles que praticavam a atividade científica.

O pesquisador, em qualquer época, sempre sentiu a necessidade de comunicar o avanço do conhecimento obtido, de trocar experiências e validar seus experimentos. A literatura registra os primórdios da comunicação, quando a troca de correspondência por meio de longas cartas entre os cientistas ou as reuniões entre eles eram uma prática comum. Posteriormente, apareceram os primeiros periódicos e sucessivamente surgiram os denominados canais *formais* de comunicação. Porém, os canais *informais* sempre existiram e permanecem incólumes à passagem do tempo e ao avanço da tecnologia. Foi o permanente desejo dos pesquisadores de divulgarem seus trabalhos de forma rápida e informal que instigou a publicação cinzenta (POBLACIÓN, 1992; CARVALHO, 2001).

Quem produz e quais são os itens que se caracterizam como literatura cinzenta?

As instituições que a publicam são variadas, como: governamentais, acadêmicas, comissões técnicas, ONGs, associações de profissionais, tanto em forma impressa como eletrônica.

Fazem parte da literatura cinzenta os trabalhos apresentados em conferências, as dissertações e teses, os boletins, os sumários, as estatísticas, os relatórios técnicos, os informes técnicos (CARVALHO, 2001).

Atualmente, acrescentam-se os correios eletrônicos, as discussões virtuais, as notícias divulgadas ou *postadas* em *blogs* e *websites*. A lista não é exaustiva, mas fornece uma ideia da modalidade de documentos que podem ser caracterizados como literatura cinzenta.

Há um volume grande de definições para esse tipo de literatura, observa-se que é mais fácil relacionar os itens que integram a LC do que defini-la (MASON, 2010). O que todos concordam é que ela difere da literatura “branca”, por não ser acessível ou adquirida pelos meios comerciais; outro ponto em comum nas definições diz respeito à dificuldade de ser encontrada, devido à falta de padronização, porém a maioria desses documentos contém dados importantes e originais (MASON, 2009).

Não procede também a ideia de que por não passarem pelo circuito comercial são menos confiáveis, isso é mera impressão, pois alguns itens da LC passam até por processos mais rigorosos de avaliação do que os da literatura branca, como é o caso das dissertações e teses.

O início da LC parece estar relacionado com a Revolução Industrial, porém foi a partir da Segunda Guerra Mundial que se observou um acréscimo significativo, devido ao grande volume de pesquisas e geração de documentos em várias áreas do conhecimento realizadas nesse período (CARVALHO, 2001).

A urgência da comunicação durante o período da guerra obrigou os pesquisadores a transmitirem seus resultados rapidamente entre os centros de pesquisa e entre as pessoas autorizadas, com a emissão de relatórios confidenciais.

O detalhamento e o progresso das pesquisas eram publicados em forma desses relatórios, conhecidos por uma codificação alfanumérica, tanto impressos como em forma de microfichas e microfilmes, e se tornaram documentos peculiares à área nuclear (CARVALHO, 2001).

Em 1946, nos Estados Unidos, o *Atomic Energy Act* criou a Comissão de Energia Atômica como sucessora do Projeto Manhattan para administrar e regular o uso da energia atômica (OFFICE OF SCIENTIFIC AND TECHNICAL INFORMATION, 2008).

Fazia parte dos programas dessa organização a liberação (*declassification*) dos dados (relatórios) obtidos por meio das pesquisas realizadas durante a Segunda Guerra Mundial e a disseminação dos documentos. Data dessa época a primeira base de dados sobre energia nuclear *Nuclear Science Data Base (NSDB)* versão impressa, a qual concentrava a quase totalidade dos documentos contendo as pesquisas sobre a área, uma obra de referência “*ancestral*” das atuais bases de dados na área nuclear.

Registrou-se este breve comentário para ressaltar que a literatura cinzenta, por coincidência, é fortemente familiar para a área nuclear campo de observação desta tese.

Foi na década de 1970 que se concentrou a preocupação com a administração da LC, e as propostas com relação à catalogação, indexação, divulgação. Apareceram também os primeiros eventos dedicados à LC, como o Seminário de York, realizado em 1978 (PEJSOVA, 2010).

Em 1970, a *Publications Board* dos Estados Unidos se transformou em *National Technical Information Service (NTIS)*, organização subordinada ao Governo americano que se tornou a maior e mais importante organização em LC. É uma fonte central de informações científicas governamentais de engenharia e de negócios, cobrindo mais de 350 áreas de conhecimento.

Em 1980 surgiu a base de dados *System for Information on Grey Literature in Europe (Sigle)*. Fruto de um consórcio de bibliotecas e centros de documentação da Europa trata-se de uma base multidisciplinar que contém relatórios, dissertações, entre outros documentos com características de LC (PEJSOVA, 2010).

O advento da Internet na década de 1990 proporcionou à LC um crescimento explosivo, mais uma vez o desejo dos envolvidos em atividades científicas de divulgarem rapidamente seus conhecimentos bem como se inteirarem do trabalho de

seus pares encontrou na Internet o meio e o instrumento ideais, minimizou-se a dificuldade de localização desses documentos, propiciou-se liberdade e rapidez na transferência de conhecimento, além de se fomentar o trabalho de equipes multidisciplinares de pesquisadores com a consequente geração de documentação dessas atividades (PEJSOVA, 2010).

Os repositórios institucionais são os maiores detentores da literatura cinza e, pela importância que estes adquiriram atualmente (ver item 3.4.1 deste trabalho, sobre repositórios), tornaram-se poderosos coadjuvantes no processo de construção do conhecimento científico bem como no processo educacional especializado (PEJSOVA, 2010).

Esse novo meio de comunicação, com todo seu alcance e domínio radical do *modus vivendi* da sociedade, potencializou o volume de literatura cinzenta. Poder-se-ia afirmar que atualmente a literatura informal domina a comunicação eletrônica, a qual exige maior rigor na seleção e identificação da procedência idônea dos documentos, principalmente para a elaboração de trabalhos científicos.

Categorizadas nesse contexto estão as dissertações e teses; conforme mencionado, dentro da cadeia de produção científica decorrente das pesquisas, esse tipo de literatura representa a culminação de um ciclo que demanda esforços, recursos e tempo dos candidatos a pesquisadores. A produção desses trabalhos está sempre vinculada a um sistema já institucionalizado; seja nas universidades, seja nos institutos de pesquisa, eles são produzidos dentro de rigorosos padrões, tanto de forma como de conteúdo, e submetidos à validação de pesquisadores seniores de reconhecido saber no assunto abordado.

No Brasil, o tratamento da LC habitualmente é realizado nas bibliotecas por meio da produção de bibliografias temáticas. Deve-se dar destaque à atuação da Dra. Dinah Aguiar Población, professora do Departamento de Biblioteconomia e Documentação da Escola de Comunicações e Artes (ECA) da Universidade de São Paulo (USP), que em 1993 criou o Núcleo de Produção Científica, cujos primeiros estudos se dedicaram à análise da literatura cinzenta produzida na áreas de Ciência da Informação, formada principalmente pelas dissertações e teses e trabalhos

científicos apresentados em eventos. Atualmente o foco de estudos deste núcleo está voltado à análise da produção dos docentes/doutores dos programas de pós-graduação em Ciência da Informação (FUNARO e NORONHA, 2006).

Na década de 1970, também época dos primeiros cursos de Pós-graduação em Ciência da informação no país, o Instituto Brasileiro de Biblioteconomia e Documentação (IBBD), atual Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), armazenava e guardava estes documentos no suplemento Livro Jornal do Brasil (CAMPELLO, 1988 citado por FUNARO e NORONHA, 2006).

Atualmente a Biblioteca Digital de Teses e Dissertações (BDTD)¹⁵, coordenada pelo IBICT em parceria com as instituições de ensino e pesquisa no país, possibilita que a comunidade brasileira da C&T deposite suas dissertações e teses. Trata-se de uma base de dados nacional que por meio de um sistema cooperativo possibilita a reunião deste tipo de material, adotando padrões de interoperabilidade consolidados em uma rede distribuída de bibliotecas digitais. Nessa rede, as bibliotecas digitais das instituições de ensino e pesquisa atuam como provedores de dados e o IBICT opera como agregador, coletando metadados de teses e dissertações dos provedores, fornecendo serviços de informação sobre esses metadados e expondo-os para coleta por outros provedores de serviço, em especial pela *Networked Digital Library of Theses and Dissertation*¹⁶ (INSTITUTO BRASILEIRO DE INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA, 2010).

Deduz-se, portanto, que esses itens de literatura se constituem em registros das linhas e tendências científicas, e se tornam um referencial para se conhecer a estrutura social da pesquisa da Instituição ou da área, em que são produzidas.

¹⁵ <<http://bdt.d.ibict.br>>.

¹⁶ ND LTD: <www.ndltd.org>.

3.4 O uso das teses como *corpus* de pesquisa

Embora não tenham sido identificados trabalhos similares a este na literatura nacional, isto é, em que tenha sido aplicada a mesma metodologia (coocorrência) e com o mesmo *corpus* de trabalho (teses), e ainda utilizando uma base de dados local, possibilitando uma comparação, estudos que utilizam as dissertações e teses como *corpus* de trabalho têm sido objeto de estudo e análise cada vez mais frequentes em nível nacional.

Os objetivos diferem na abordagem, alguns analisam o conteúdo dos trabalhos como, por exemplo, a tese de MACHADO (2007), que utilizou as teses para demonstrar a transferência do conhecimento para a sociedade, ou de NASCIMENTO (2010), que efetuou um mapeamento temático das teses com o objetivo de verificar sua adequação às áreas de concentração e linhas de pesquisa do Programa onde foram produzidas.

A técnica de coocorrência de palavras (*co-word analysis*) foi utilizada por GONSALEZ-SAES (2005), em sua tese, com o objetivo de medir a atividade científica nacional na área de saúde construindo mapas tecnológicos, com dados extraídos dos artigos registrados na WoS.

LOUREIRO- ALVES (2010), apresentou uma tese cujo objetivo era conhecer os aspectos relacionados à institucionalização da Ciência da Informação do ponto de vista cognitivo e social, para tanto utilizou as dissertações e teses depositadas na base de dados da CAPES, no período 1985-2007 e as técnicas bibliométricas para análise.

Em nível internacional as dissertações e teses também são objeto de estudo, sob os mais diversos aspectos, mas não no que se refere ao uso da técnica de coocorrência de palavras para verificar a produtividade destas, no entanto encontram-se diversos estudos bibliométricos envolvendo as teses na Espanha, uma

vez que esse país dispõe de uma base de dados nacional, digital e especializada reunindo as dissertações e teses (TESEO)¹⁷.

Cita-se como exemplo a tese de VALLEJO RUIZ (2005), a qual faz uma análise longitudinal objetivando verificar o nível de expansão da área da Matemática, utilizando técnicas de bibliometria.

3.5 As bases de dados (BDs)¹⁸

Consideram-se importantes os comentários a respeito do conceito histórico e do funcionamento das bases de dados, uma vez que um dos objetivos específicos deste trabalho diz respeito ao papel multifuncional das bases de dados bibliográficas, no caso como provedoras de matéria-prima para os estudos bibliométricos. As bases de dados se inserem no contexto da organização da informação, representação descritiva e ciência da computação

Uma base de dados é um conjunto de informações armazenadas em suportes legíveis por computador, organizadas por registros formados por campos que permitem recuperar todos os tipos de informação (referências, textos completos, imagens, dados numéricos, e assim por diante) (YUNTA, 2001).

As bases de dados são elaboradas com o objetivo de fornecer informação atualizada, precisa e confiável atendendo a uma demanda específica. CENDÓN (2002), registrou a seguinte classificação para as bases de dados quanto a modalidade:

- ✓ bases de dados bibliográficas ou referenciais onde cada registro corresponde a um documento que permite ao usuário localizar determinadas publicações como: artigos de periódicos, teses, livros, relatórios de pesquisa, artigos apresentados em eventos, uma vez que fornecem todos os dados dos documentos, autor, título, ano, local de publicação, e algumas também apresentam resumo;

¹⁷ Base de dados nacional que reúne todas as teses defendidas em universidades espanholas desde 1976 (www.mcu.es/TESEO). Nos Estados Unidos se encontra uma das maiores bases de dados dedicadas a reunir teses como o *Proquest Digital Dissertations* <http://lib.umi.com>

¹⁸ Neste trabalho, a abordagem para base de dados é sempre do ponto de vista documental ou bibliográfico.

- ✓ bases de dados fatuais (ou numérica) apresentam informações de caráter numérico como dados estatísticos, séries históricas, resultados de pesquisa, índices de inflação e outras informações de teor similar;
- ✓ bases de dados com texto completo, como o nome indica, apresentam o documento em texto completo e não apenas a referência. Atualmente esta modalidade é a mais disponibilizada, uma vez que o acesso ao documento é imediato.

YUNTA (2001), acrescenta ainda outra modalidade de base de dados denominada diretórios; esta modalidade apresenta dados sobre pessoas ou instituições especializadas em determinada atividade como diretórios de profissionais, de empresas de bibliotecas, e assim por diante.

Dependendo do tipo de cobertura temática, também podem ser multidisciplinares, isto é, abrangem vários assuntos simultaneamente, um exemplo disto seria a base internacional *Web of Science ou Scopus*, e especializadas, que focam em determinado assunto, por exemplo, a base de dados INIS, especializada na área nuclear e áreas relacionadas. Quanto à abrangência geográfica, podem ser locais, por exemplo, a base de dados da Produção Científica do IPEN (PTC) desenvolvida para atender a uma comunidade específica do Instituto; nacionais, abrangem a literatura de determinado país, por exemplo, a Base de Dissertações e Teses gerenciada pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), e internacionais, como por exemplo a MEDLINE, produzida pela *National Library of Medicine* dedicada à área de saúde em nível internacional.

Inicialmente, as bases de dados, quando surgiram, nas décadas de 1960 e 70, eram produzidas quase exclusivamente por órgãos governamentais, porém essa posição foi mudando e atualmente a situação é inversa, as empresas privadas dominam o mercado de produção de bases de dados, embora ainda existam as BDs produzidas por órgãos governamentais (NEUFELD e CORNOG, 1986).

Há alguns atributos que são fatores de sucesso para as bases de dados como, por exemplo, a atualização e disponibilidade de consulta *on-line*, porém dentro dessas características, a estrutura adotada para a alimentação, organização e

recuperação da informação é um ponto crucial que está diretamente relacionado com o desempenho e a qualidade das bases (YUNTA, 2001).

Quanto mais estruturado e organizado estiver o conteúdo da base, melhor será o seu desempenho e, conseqüentemente, maior será sua aceitação por parte dos usuários.

Historicamente, as bases de dados, tal como são conhecidas hoje, nasceram do desenvolvimento da automação e da necessidade de organizar, sistematizar e permitir a consulta ao grande volume de informação existente na época, o período pós-guerra, a década de 1950 (NEUFELD e CORNOG, 1986).

Não se tem registro exato de quando surgiu a primeira base de dados, desde que qualquer arquivo que fosse lido pelo computador se denominava *base de dados*. Também não é demais lembrar que *arquivos lidos pelo computador* é força de expressão, uma vez que se faz referência aos estoques de cartões perfurados usados para serem lidos pelos primeiros computadores. Utiliza-se aqui o termo *base de dados* para simplificar a leitura, porém essa denominação está fortemente associada ao uso do computador (NEUFELD e CORNOG, 1986).

As tentativas anteriores (ao aparecimento do computador) de se armazenar e processar o conhecimento (informação) existente na época falharam, devido, principalmente à incapacidade de manusear grandes volumes de informação, ou seja, devido à inviabilidade tecnológica. Por isso, o computador foi a mola propulsora para o desenvolvimento da indústria da informação, e mais especificamente para as bases de dados. A capacidade de armazenar grandes quantidades de informação em fitas magnéticas, e posteriormente em discos, permitiu o desenvolvimento da pesquisa *on-line*, a qual realmente revolucionou a forma de acessar a informação (NEUFELD e CORNOG, 1986).

Consta na literatura que uma das primeiras bases de dados voltadas para a pesquisa bibliográfica foi apresentada pela *National Library of Medicine* em 1964 e iniciou o serviço de recuperação da informação da literatura médica, o MEDLARS (atual MEDLINE) (BUXTON, 1992).

Na década de 1980 houve uma explosão de bases de dados, muitas empresas comerciais se interessaram pelo negócio da informação percebendo a grande procura e também porque o avanço da tecnologia viabilizava o comércio da informação. O número de BDs passou de 300 em 1975 para 800 em 1980 e para 2.400 em 1984, incluem-se nesse número as bases de dados numéricas bem como as BDs de texto completo que já começavam a despontar, embora em número reduzido (NEUFELD e CORNOG, 1986).

Um dos últimos levantamentos oficiais sobre o cenário de bases de dados realizado por WILLIAMS (2002), em um artigo introdutório na *Gale directory of Databases*, e já na época da Internet, registrou que em 2001 havia 12.900 bases de dados, com 16.880 milhões de registros, de conteúdos variados, porém predominantemente bibliográficas.

Segundo TENOPIR (2003), a grande disseminação e aceitação dos microcomputadores modificaram também o ambiente das bases de dados. Da metade da década de 1980 até a metade da década de 1990, aproximadamente, houve grande ênfase no uso da versão das bases de dados em CD-ROM. Em 1993, novamente o ambiente da informação foi drasticamente modificado pelo uso da Web, que superou o uso deste suporte (CD-ROM); nos anos subsequentes cada base de dados disponível *on-line* disponibilizou suas interfaces de consulta compatíveis com o ambiente Web.

A rapidez com que uma tecnologia supera outra vai tornando obsoletos dispositivos e serviços em espaços de tempo cada vez menores.

A capacidade de armazenamento e recursos de mídia logo ofereceram as bases de dados com acesso ao texto completo. O surgimento dos *e-journals*, *e-articles* e *e-books* causou um grande impacto no mundo da informação. Os usuários das bases de dados atualmente têm expectativa de acesso ao texto completo ao realizarem as buscas bibliográficas. As bases de dados sem *links* estão com os dias contados no mundo do acesso ao texto completo (TENOPIR, 2003).

Segundo TENOPIR (2001), atualmente a figura de base de dados única tende a desaparecer, os serviços de informação modificaram os serviços prestados pelas

bases de dados, a oferta foi ampliada pelas *megabases* de dados, como por exemplo, a *ProQuest*, *Dialog*, *Lexis-Nexis*. Segundo declarações do gerente de uma *megabase* de dados a *OneFile*, HACKWOOD (2001) citado por TENOPIR (2001), os usuários veem a Internet como uma fonte inesgotável de informação e as bibliotecas têm que disputar com ela a atenção dos usuários e prossegue: o modelo tradicional de acesso e venda foi definido na época do CD-ROM (quando era habitual se adquirir uma só base de dados), atualmente este mercado está dominado pelas subscrições a bases de dados *on-line*, técnica e financeiramente não há mais necessidade de segmentar o acesso às bases de dados.

Em 1998, WORMELL publicou um artigo com o objetivo de chamar a atenção para a utilidade das bases de dados não somente como instrumentos de recuperação de informação, mas como instrumentos de análise apropriados para identificar as tendências e a atuação da sociedade, as disciplinas científicas e das áreas de produção e consumo. Comentou também a oportunidade dos profissionais da informação de agregar valor às funcionalidades das bases de dados, embora ainda tivessem pouca compreensão dessa potencialidade.

MOED (1988); STEFANIAK (1987); INGWERSEN e WORMEL (1990), HOOD e WILSON (2003) são autores que também comentaram sobre a utilização das bases de dados como ferramentas importantes no sentido de fornecerem a matéria-prima (informação) para a aplicação de métodos bibliométricos e produção de estudos métricos da ciência, embora todos sejam unânimes em apontar como principal ponto negativo a qualidade dos registros, uma vez que a bibliometria requer fundamentalmente a exatidão dos dados sob pena de inviabilizar ou comprometer todas as análises. Também reconhecem que as bases de dados não foram criadas com essa finalidade, embora a qualidade dos dados devesse estar presente independentemente dos objetivos finais.

PEREIRA (1999), a esse respeito comenta que as bases de dados são fundamentais para a tarefa daqueles que desejam analisar a Ciência e Tecnologia e suas relações como o desenvolvimento social e econômico, neste aspecto é importante a qualidade dos dados que subsidiam tais análises. Este aspecto merece

atenção especial, já que é fundamental para a qualidade das bases de dados o cuidado com o conteúdo que flui nas estruturas de armazenamento, dito de outra forma, uma base de dados com erros de conteúdo pode comprometer um trabalho de análise e acarretar prejuízos financeiros. A autora prossegue afirmando que estabelecer uma política para aferir a qualidade das bases de forma contínua contribuirá para uma melhoria das bases.

Na área científica grande parte dos estudos bibliométricos ou cientométricos utiliza os dados das principais bases de dados internacionais, como a *Web of Science (WoS)* ou, *Scopus*, esses sistemas de informação já disponibilizam recursos e instrumentos para medição, possibilitam, por exemplo, fazer o *download* da informação selecionada para softwares mais especializados, porém nem sempre os dados estão livres de inconsistências, além de essas bases internacionais serem parciais, no sentido de não incluírem a literatura produzida em todos os países de forma exaustiva, isto é, a informação obtida necessita de análise, correção, padronização, enfim, passar por um processo de consistência dos dados, antes de ser trabalhada. De qualquer forma, ainda assim são as principais fontes utilizadas para o desenvolvimento dos estudos métricos (KOBASHI, 2004).

No país não há disponibilidade de bases de dados sobre a produção científica a exemplo das internacionais já citadas. É fato reconhecido pelos profissionais da informação que o país carece de bases de dados nacionais abrangentes, organizadas e permanentes, isto é, que possam fornecer dados consistentes em um longo período de tempo, porque para realizar estudos de mensuração da ciência é necessário dispor de séries temporais. Nesse sentido, o comentário de KOBASHI (2004) é oportuno. Diz a professora, pesquisadora com autoridade que lhe confere sua grande experiência nesse campo de atuação e após estudo realizado para o desenvolvimento de um projeto de pesquisa sobre o tema “desenvolvimento de indicadores da produção científica da área de Ciência da Informação”, realizado dentro do Programa de pós-graduação em colaboração entre a USP e a PUC de Campinas, São Paulo:

A precariedade das bases de dados sobre a produção científica brasileira fica evidente na pesquisa. É alarmante o grau de desorganização, dispersão e falta de padronização dos repositórios de produção científica nacionais. Esses fatos indicam que a realização de análises rigorosas, qualquer que seja a área de conhecimento considerada, encontra barreiras muitas vezes intransponíveis (KOBASHI, 2004).

Se em nível nacional há ainda esta lacuna, no entanto, já se encontram no país, ainda que de forma reduzida, iniciativas institucionais, locais e temáticas. A título de exemplo, registra-se o excelente trabalho desenvolvido pela *Scientific Electronic Library* (SciELO) da Bireme. Mediante parceria entre a Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) e o Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (Bireme), criou-se o projeto que passou a receber o nome de SciELO. Ele visa, em última análise, tornar mais visível a literatura científica brasileira e, ao mesmo tempo, criar uma base de dados que possa ser utilizada para a obtenção de informações úteis em termos de sociologia da ciência no Brasil (MENEZHINI, 1998).

Trata-se de um modelo para a publicação eletrônica cooperativa de periódicos científicos via Internet, além da organização de bases de dados bibliográficas e de textos completos, recuperação de textos por seu conteúdo, preservação de arquivos eletrônicos e a produção de indicadores estatísticos de uso e impacto da literatura científica. A Metodologia também inclui critérios de avaliação de revistas, com base nos padrões internacionais de comunicação científica. Os textos completos são enriquecidos dinamicamente com links de hipertexto com bases de dados nacionais e internacionais (SciELO, 2010).

São bons exemplos, também, as universidades e os institutos de pesquisa, os quais mantêm bases de dados institucionais ou temáticas que congregam toda a produção científica da Instituição de origem, embora sempre com a finalidade de armazenar e permitir a recuperação dos dados.

Pela experiência da autora no gerenciamento de bases de dados bibliográficos científicos, para serem utilizadas em estudos bibliométricos, todas, em maior ou

menor grau, demandariam esforço expressivo de complementação, padronização e consistência dos dados, isto é, reformatação dos dados para aplicar as técnicas bibliométricas.

Cita-se como exemplo a base de dados do Instituto Nacional de Pesquisas Espaciais (INPE), a base de dados da Produção Científica Cultural e Artística da Universidade de São Paulo (USP), a da Empresa Brasileira de Produtos Agropecuários (EMBRAPA), ou ainda a própria base de Produção Científica do IPEN, objeto de estudo deste trabalho, entre outras.

Na opinião de YUNTA (1998), as bases de dados são produtos vivos que exigem esforço contínuo de alimentação (*input*), manutenção, controle de saída (*output*) e desempenho, exigem também a atuação de equipes multidisciplinares, altos investimentos e avaliação constante para garantir a qualidade do produto.

O que se quer demonstrar é que desde a década passada já existe a percepção de que as bases de dados têm-se constituído em fornecedoras de matéria-prima para a elaboração de estudos métricos, viabilizando a produção de indicadores científicos, tornando-se instrumentos multidisciplinares, no entanto, permanece ainda, em todas elas, a questão de padronização e consistência dos metadados.

Na opinião da autora deste trabalho, partindo da experiência com o trabalho com bases de dados, o processo de construção e disponibilização destes dispositivos institucionais de recuperação da informação evoluiu e continuará evoluindo para atender a novas formas e necessidades de armazenamento e tratamento da informação, no entanto, seu futuro dependerá em grande parte da capacidade de esses sistemas utilizarem padrões de dados reconhecidos por todas as mídias eletrônicas com acesso à Internet, bem como incorporarem em seu conteúdo diferentes suportes como, por exemplo, imagens em 3D e áudio e formas de acesso.

As bases de dados na área médica são um bom exemplo desse avanço. Elas se interligam e permitem importação e exportação de dados; é possível trabalhar

simultaneamente com várias BDs em tempo real de forma simples e principalmente segura.

Concorda-se com a opinião de TENOPIR (2001), quando afirma que as bases de dados referenciais estão se tornando obsoletas, e neste sentido registra-se a percepção e esforço de adaptação à demanda dos usuários, de algumas bases, como por exemplo, a base internacional INIS, que a partir de 2009 tornou o acesso livre à base e se encontra atualmente em estágio híbrido de operação, isto é, com 3,3 milhões de referências bibliográficas de literatura nuclear, oferece acesso ao texto completo a 280 mil itens utilizando a tecnologia Google (INTERNATIONAL NUCLEAR INFORMATION SYSTEM, 2010).

3.5.1 Os repositórios de informação

O avanço das inovações tecnológicas, principalmente aquelas ligadas à informação e à comunicação científica, promove mudanças constantes e cada vez mais velozes; na era da comunicação eletrônica o conceito de *antigamente* pode ser tanto o ano passado como a semana passada, dada a velocidade com que elas acontecem. Cada mudança tecnológica promove nova mudança no cenário da informação. Diante desses avanços, surge a necessidade do desenvolvimento de ferramentas para a otimização dos processos de organização, descrição, recuperação, intercâmbio e preservação digital de informações.

A exemplo de outras áreas, a área de informação científica também se beneficiou expressivamente com o avanço e a consolidação do uso da Internet. É oportuno lembrar as palavras de MEADOWS (1999), que ainda na década de 1990 já comentava:

A comunicação científica sempre passou por mudanças. Quando os principais veículos eram a palavra escrita e impressa, isso acontecia com relativa lentidão. Os novos meios de comunicação criados pela tecnologia da informação aceleram este ritmo. A comunicação científica passa agora por um período de rápida evolução. A questão básica continua sendo como as propriedades de um dado

veículo podem ser mais bem utilizadas para atender às necessidades de comunicação da comunidade científica (MEADOWS, 1999 p. 245).

As poucas iniciativas de construção e manutenção de BDs já consolidadas no país, nas instituições já citadas, estão atentas ao desenvolvimento das suas áreas, ao avanço das tecnologias de informação, e principalmente aos movimentos reivindicatórios da área acadêmica em nível internacional, como prova disso já ampliaram o escopo de serviços, atualizaram as ferramentas de tratamento da informação e estão se tornando bases de dados digitais, inseridas no conceito de bibliotecas digitais.

As bibliotecas digitais são consideradas sistemas de informação complexos que abrangem a criação, gestão, distribuição e preservação de informação, por meio digital, para permitirem uma interação eficaz e eficiente entre as diversas comunidades que se beneficiam do conhecimento disponibilizado nesses ambientes (TAMMARO; SALARELLI, 2006).

Foi neste contexto, em agosto de 1991, que apareceu o primeiro repositório de informação, com documentos eletrônicos baseados na filosofia dos arquivos abertos (*open archives*). Essa iniciativa foi pioneira na reflexão sobre o processo de publicação e reorganização das publicações científicas, bem como de suas influências na comunicação científica eletrônica (BORGES, 2008).

Trata-se de outra modalidade de organização, descrição, recuperação, disseminação e armazenamento da informação digital, os denominados repositórios na essência são grandes bases de dados cuja filosofia e logística de funcionamento manutenção e consulta incorporam toda a tecnologia de *software* e *hardware* e comunicação disponível até o momento (KURAMOTO, 2009).

A filosofia de funcionamento dos repositórios é inovadora e quebra paradigmas consolidados na área acadêmica, trata-se de um modelo democrático e eficiente para disseminação dos resultados das pesquisas assegurando a transparência e o acesso universal à produção científica sem custo, daí a denominação “arquivos abertos”, por meio da formação de bases de dados como repositórios de autoarquivamento em que os pesquisadores depositam seus artigos

publicados e aceitos em revistas científicas, ou repositórios de dissertações e teses produzidas por uma Instituição, ou incentiva ainda que grupos ou projetos de pesquisa criem suas próprias bases de dados contendo materiais como relatórios de pesquisa, relatos de experimentos, artigos apresentados em eventos, e assim por diante (BORGES, 2008).

Se o conceito de arquivos abertos (*open access*) é inovador, a logística operacional que envolve a criação e manutenção dos repositórios também é notavelmente inovadora. O autoarquivamento (*self archiving*) diz respeito ao direito do próprio autor de depositar seu texto para publicação em um repositório de arquivos garantindo visibilidade a seus trabalhos de pesquisa sem a interferência de terceiros, isto é, libertando-se do monopólio das grandes editoras científicas que até então detinham os direitos da propriedade intelectual e editorial da publicação.

O movimento de arquivos abertos tem como referência a Convenção de Santa Fé realizada em 1999 no Novo México, organizada por Paul Ginsparg, Rick Luce e Herber Van Sompel, físicos e acadêmicos atuantes nos Estados Unidos (BORGES, 2008).

Durante essa Convenção, foram definidos os princípios básicos de uma nova era para publicação científica, e a disseminação de resultados de pesquisas. Foram discutidos o funcionamento dos serviços de autoarquivamento, o estabelecimento de normas de adesão à iniciativa e à definição de metadados, especificações de protocolo e padronização de linguagens. Foi definido também um padrão para troca de informações entre arquivos eletrônicos, que ficou conhecido como *Open Archives Initiative (OAI)*, o qual objetiva a integração de arquivos *e-prints*, promovendo a interoperabilidade e o intercâmbio automático de metadados entre bibliotecas digitais KURAMOTO (2009); BORGES (2008); MUELLER (2006); SUAIDEN (2006).

Vista por outro ângulo, e sintetizando, a proposta de criação de repositórios de informação aliada ao conceito de arquivos abertos é uma reação da comunidade acadêmica:

- ✓ ao modelo de negócios praticado pelas editoras de revistas científicas, ou seja, o custo cada vez mais alto das assinaturas de revistas, inversamente

proporcional aos orçamentos das bibliotecas, cada vez mais reduzidos, inviabilizando a manutenção das coleções;

- ✓ à demora na publicação dos *papers* devido à burocracia envolvida;
- ✓ à rigidez nos critérios adotados pelos comitês editoriais, privilegiando determinados temas e marginalizando outros;
- ✓ à imposição dos editores obrigando os autores a abdicarem dos seus direitos editoriais e intelectuais;
- ✓ ao atual estágio de desenvolvimento da tecnologia da informação, que permite mais velocidade e promove novas formas de produção, preservação, disseminação e, principalmente, a comunicação entre pares (BORGES, 2008).

A incorporação dos mecanismos de interoperabilidade propiciada pelo uso dos arquivos abertos tornou possível a criação de grandes redes de bibliotecas digitais. Excelente exemplo é o projeto da biblioteca digital em rede da Universidade de Virgínia. O projeto *Network University Digital Library* (NUDL) foi desenvolvido para que a propriedade intelectual produzida nas universidades pudesse ser mais acessível, estimulando a transferência de tecnologia, a colaboração internacional e o conhecimento compartilhado de todas as disciplinas. O NUDL permite a configuração de uma rede de bibliotecas digitais de teses e dissertações por meio do *Network Digital Library of Theses and Dissertations* (NDLTD), uma iniciativa internacional para melhorar a educação, permitindo que os estudantes publiquem suas teses e dissertações em meio eletrônico via Web através de um procedimento de submissão, utilizando o protocolo OAI (BORGES, 2008).

No Brasil, conforme já comentado, destaca-se a iniciativa do Instituto Brasileiro de Informações em Ciência e Tecnologia (IBICT), que coordena o projeto da Biblioteca Digital de Teses e Dissertações (BDTD) como uma das notáveis iniciativas, em nível nacional, de desenvolvimento de uma base de dados digital, cujo objetivo é integrar os sistemas de informação de teses e dissertações existentes nas Instituições de Ensino Superior (IES) brasileiras, além de estimular o registro e a publicação de teses e dissertações em meio eletrônico. A BDTD utiliza as

tecnologias do *Open Archives Initiative (OAI)* e adota o modelo baseado em padrões de interoperabilidade consolidado em uma rede distribuída de bibliotecas digitais de teses e dissertações. O papel do IBICT na rede é de agregador, participando como coletor, provedor de dados e provedor de serviços, ou seja, percebe-se que é um modelo diferenciado de construção da base de dados digital, embora destinado a reunir um só tipo de material, qual seja, as dissertações e teses (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2010).

A BDTD é uma das ações do programa que contam com o apoio financeiro da Financiadora de Estudos e Projetos (FINEP) e em uma ação finalística integra-se ao ND LTD (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2010).

Na realidade, o novo ambiente de construção de repositórios desafia os sistemas tradicionais de comunicação e divulgação científica, que até hoje ofereceram acesso restrito, visto que os artigos publicados em revistas e anais de congressos normalmente estão sob a guarda de editoras comerciais. A distribuição e circulação de publicações, como teses, dissertações, relatórios de pesquisa e experimentos era bastante limitada.

A adesão aos repositórios institucionais é visível pelo número de signatários da Declaração de Berlim¹⁹, exprimindo, assim, seu compromisso com o livre acesso à informação. O país mais avançado no estabelecimento de repositórios institucionais é o Reino Unido (BORGES, 2008).

Trata-se do início de uma revolução cujos limites ainda não podem ser previstos, mas que encerra em si a promessa de democratização e acesso universal aos recursos de informação. Os repositórios propiciam uma “reforma” no processo da comunicação científica, não somente democratizando o acesso aos resultados das pesquisas, mas também proporcionando à comunidade acadêmica a possibilidade de reassumir o controle da publicação científica.

¹⁹ Manifesto emitido, em favor do livre acesso à publicação científica, durante o congresso organizado pelo Max Plank Institute e pela European Cultural Heritage Online, realizado em Berlim em 2003. O manifesto na íntegra está acessível em: <<http://oa.mpg.de/lang/en-uk/berlin-prozess/berlin-konferenzen>>. Acesso em jan. 2011.

A utilização da tecnologia da informação não se resume ao uso extensivo da tecnologia digital, mais do que isso, modifica os métodos de trabalho e representação do conhecimento e motiva uma redefinição nos papéis dos diversos atores que participam no ciclo da informação, comunidades científicas, editores, bibliotecas, profissionais da informação, entre outros (BORGES, 2008).

3.5.2 Os sistemas de recuperação da informação

Ao se abordar o tema de recuperação da informação, habitualmente conhecida dentro da Ciência da Informação como RI, faz-se necessário contextualizar o termo.

A recuperação da informação (RI) é uma das etapas dentro de um conceito maior, qual seja, a organização da informação, e para entender um pouco a respeito deste conceito também muito amplo cita-se SVENONIUS (2001), segundo a qual organizar a informação objetiva agrupar informações por semelhança, e está relacionado ao documento impresso, à sua ordenação e ao arranjo para torná-lo disponível, ou, dito de outra forma, organizar para poder recuperar ou encontrar a informação mais rapidamente, para tanto utiliza-se procedimentos reconhecidos pelo tratamento da informação, a partir de metodologias formais e consolidadas em que não se trata de alterar o documento, porém se trata de criar novas formas de representá-lo, viabilizando seu fluxo e uso.

O termo recuperação da informação é contemporâneo ao aparecimento do uso do computador e a literatura a respeito é extensa. KENT (1972), faz referência ao termo comentando que diferentemente da ideia de se recuperar qualquer outro objeto, não significa procurar algo perdido, e sim trata-se da busca de algo que foi anteriormente organizado. Em uma versão mais atual, INGWERSEN (1992), entende que a RI abrange processos de representação, armazenamento, busca e encontro de informação relevante; definindo como relevante, o grau de correspondência ou utilidade entre o documento localizado e a questão (necessidade) de informação de uma pessoa.

Tão importante quanto as definições sobre RI são os modelos utilizados, estes podem ser modelos clássicos e modelos estruturados. No modelo clássico, cada documento é descrito por um conjunto de palavras-chave, onde se representa o documento tematicamente, o processo é denominado indexação (é o caso dos procedimentos utilizados na base de dados que serviu de fonte fornecedora da matéria-prima para realização deste trabalho, a PTC,). Nos modelos estruturados especificam-se, além das palavras-chave, outras informações a respeito da estrutura do texto, como coeficientes de proximidade, parágrafos do documento, e assim por diante (ROCHA, 2006).

O processo de indexação se insere em um macrouniverso, o da análise documentaria, área teórica e metodológica cujo objetivo é o tratamento temático do documento (GARDIN, 1981 citado por SILVA e FUJITA, 2004). A indexação é então a etapa da representação, fase final de análise documentária, para tanto se utilizam as linguagens documentarias. LANCASTER (1993), registrou que a indexação pode ser seletiva, constituída de descritores que representem de forma genérica o assunto do documento, ou mais exaustiva, com descritores que representem assuntos mais específicos.

Na etapa da análise documentária o indexador efetua a leitura documentária para identificar os conceitos que compõem o tema do documento; segundo SILVA e FUJITA (2004), a leitura documentária é o momento em que o indexador realiza a análise do assunto e identifica os conceitos importantes tratados nos documentos a fim de representá-lo. É aqui que os aspectos lógicos, linguísticos e cognitivos envolvidos na indexação representam fatores de interferência por parte do indexador.

Este aspecto da indexação é especialmente importante para a compreensão deste trabalho, uma vez que a técnica de análise de coocorrência utilizou as palavras-chave atribuídas pelo indexador como objeto de correlação entre os artigos e as teses. CALLON et al. (1991) e WHITTAKER (1989) registraram esta etapa como uma limitação na utilização dos estudos de correlação que utilizam os descritores de um vocabulário controlado, e o que eles denominaram *efeito/indexador* seria

exatamente na etapa de atribuição dos conceitos onde a subjetividade do indexador pode interferir (ver p. 139 deste trabalho).

Na indexação realizada manualmente, isto é, realizada por profissionais da informação, cabe a estes atribuir os conceitos que representam os documentos e respondem à busca do usuário, para isso são utilizados os Tesauros, ou vocabulários controlados específicos de cada área de conhecimento.

Conforme se pode perceber, a indexação é tarefa complexa que requer o trabalho de equipes multidisciplinares, exige do indexador tanto a combinação de conhecimentos técnico-teóricos (linguagens documentárias) como grande familiaridade com a área indexada, para representar os documentos por meio de conceitos, retirados do vocabulário controlado, os quais se tornarão pontos de acesso para a recuperação da informação por parte dos usuários, de onde se deduz que uma indexação inadequada compromete todo o processo de recuperação da informação.

Um componente igualmente importante dentro do processo de indexação são os Tesauros, embora não sejam os únicos instrumentos. SVNENONIUS (1986), analisou os instrumentos utilizados para a indexação, estes são: as classificações, os cabeçalhos de assunto e os Tesauros, estes últimos serão abordados por interessarem ao desenvolvimento deste trabalho.

A principal característica dos Tesauros é o arranjo hierárquico e as relações semânticas entre os termos. Um tesouro reúne a terminologia de uma área, é um conjunto de unidades ou termos que representam um conhecimento especializado.

Os Tesauros ou vocabulários controlados contribuem para ampliação dos resultados relevantes desejados pelos usuários, eles são produzidos para aumentar a precisão no resultado das pesquisas. O controle é exercido de várias formas: hierárquica, relacional, ortográfica, por exemplo, o controle da sinonímia (variações ortográficas equivalentes para a mesma palavra) e da homonímia (variações léxicas equivalentes) (SVENONIUS, 1986).

Observa-se que esse instrumento utilizado organiza e viabiliza o acesso aos itens de informação ao auxiliarem nas tarefas de gestão lógica dos documentos e sua representação (indexação e descrição).

A estratégia de busca é outra etapa importante na recuperação da informação e no uso do vocabulário controlado. LOPES (2002), registra que a estratégia de busca pode ser definida como uma técnica ou conjunto de regras para possibilitar o encontro entre uma pergunta formulada e a informação armazenada em uma base de dados, isto é, a estratégia de busca irá determinar a eficácia na obtenção dos resultados, porém tanto a indexação como a recuperação são fortemente influenciadas pela qualidade da linguagem documentária da área da qual o sistema trata, bem como pela política estabelecida para o funcionamento do sistema de informação.

O que se observa é que há atualmente diversas frentes de pesquisa para tratar do desenvolvimento da recuperação da informação, principalmente depois do uso massivo dos serviços de busca da informação (Google, Yahoo e outros), e onde um dos principais problemas da recuperação é a seleção de quais documentos são relevantes e quais podem ser descartados. A tarefa de seleção em sistemas automatizados é executada automaticamente por meio de algum tipo de algoritmo que baseado em heurísticas previamente definidas decide quais são os documentos relevantes (ROCHA, 2006), porém não são considerados sistemas eficientes.

A título de exemplo também cita-se vertente da Web semântica (nome genérico do projeto em desenvolvimento que, dirigido pelo consórcio *World Wide Web*, pretende atribuir inteligência e contexto aos códigos *XML*²⁰ utilizados para a confecção das páginas Web). Trata-se de um segmento complexo que demandaria a integração de pesquisa em diversas áreas de conhecimento, como linguística, psicologia cognitiva, ciência da computação, sociologia, entre outras (ROCHA, 2006).

²⁰ XML, ou *eXtended Markup Language*, é uma linguagem de programação, um padrão para a formatação de dados, ou seja, uma maneira simples de organizar as informações de determinado documento para ser lido pelo computador.

3.5.3 A base de dados da Produção Científica do IPEN (PTC)

A base de dados da Produção Científica do IPEN (PTC) conta atualmente com 15.200 mil documentos, todos representados tematicamente; nessa base, aproximadamente 60% dos itens está disponível em texto completo, o restante encontra-se em processo de digitalização.

Foi criada em 1974, construída no padrão internacional da base de dados do INIS, com o principal objetivo de reunir de forma diferenciada toda a produção científica da Instituição, dissertações e teses, artigos publicados em periódicos e em eventos tanto nacionais como internacionais relatórios técnico-científicos, livros e capítulos de livros.

Com o decorrer do tempo, atribuíram-se outras funções finalísticas à base. Atualmente, a base PTC é provedora de informação para: **a)** a base de dados INIS, especializada na área nuclear, que por sua vez fornece informações para a formação de outra base de dados americana a *Energy Data Base* (EDB); **b)** a produção científica da USP de dissertações e teses; **c)** o fornecimento de dados estatísticos anuais para a área de planejamento do IPEN; **d)** a consulta dos alunos de pós-graduação em início de trabalho e familiarização com a área; **e)** a preservação da memória científica da Instituição.

3.5.4 Os procedimentos utilizados para a indexação dos documentos

Para se alimentar a base de dados, utilizam-se os próprios documentos produzidos pelos pesquisadores do IPEN, os quais são encaminhados periodicamente para a biblioteca, inicialmente anexados a um formulário disponível na intranet. Atualmente, dispõe-se de uma interface, *on-line*, via intranet da Instituição, acessível a todos os pesquisadores do IPEN, em fluxo contínuo, a qual permite a coleta automática e o autodepósito do documento na versão texto completo. Uma vez coletado o material, a equipe do setor de documentação valida o recebimento e o nível monográfico do documento, extrai os metadados, indexa-os

utilizando o Tesouro INIS, para finalmente serem acrescentados à base PTC. A coleta automática dos documentos também fornece dados para subsidiar o Sistema Gerencial de Planejamento do Instituto (SIGEPI), cumprindo assim outra função institucional.

A FIG. 8 apresenta de forma simplificada a logística de funcionamento da coleta da produção científica dentro da Instituição IPEN. Observa-se que se trata de um modelo de coleta híbrido, entre o funcionamento de uma base de dados tradicional, porém, com o conceito de autoarquivamento de um repositório institucional.

As teses, especificamente, são encaminhadas à biblioteca tanto em forma impressa como digital. Elas cumprem, também, outra rotina, fruto de acordos institucionais: são depositadas primeiro em sistemas de informação tanto nacionais, como o BDTD do IBICT via USP, como internacionais, como o *INIS da International Atomic Energy Agency* (IAEA), em Viena.

A principal ferramenta de indexação é o Tesouro produzido pelo INIS desde 1970 em forma impressa, atualizado anualmente, e sua consulta está disponível *on-line*. O arranjo do Tesouro obedece a uma ordem hierárquica, é um instrumento que garante aos profissionais da informação e aos pesquisadores a padronização da linguagem da área nuclear e ciências relacionadas. O uso do Tesouro, a exemplo da base INIS, é um dos mecanismos que garantem a qualidade da base PTC.

Os descritores (conceitos) são apresentados em ordem alfabética e possibilitam a localização tanto de termos gerais como de termos específicos. Além disso, a exemplo da maioria dos Tesouros, apresentam notas explicativas para termos que possam apresentar ambiguidades e frequência de utilização; há indicação de termos mais abrangentes, menos abrangentes, relacionados, obsoletos ou menos apropriados para determinado conceito, direcionando para um termo mais adequado.

O INIS adota um sistema descentralizado de alimentação da base de dados, isto é, os 150 países membros da *International Atomic Energy Agency* (IAEA) coletam e processam seu próprio material bibliográfico e o enviam para Viena

(Áustria), para se confeccionar a base de dados internacional, na área nuclear, o INIS, as normas para indexação são bastante detalhadas e padronizadas.

Na área nuclear no país, a indexação é realizada por especialistas, em geral pesquisadores; porém, os próprios profissionais da informação que trabalham na biblioteca do IPEN estão aptos a efetuar a indexação após um período de treinamento viabilizado por meio de cursos realizados em Viena, na sede do INIS, ou também no Centro de Informação Nuclear (CIN/CNEN) no Rio de Janeiro, tanto na filosofia e na operacionalidade do Tesauro como no sistema de recuperação e funcionamento da base de dados do INIS. O número médio de palavras-chave sugerido pelo INIS é de oito, porém, cabe ao indexador decidir sobre o número adequado a ser atribuído a cada documento dependendo da profundidade e da abrangência do tema tratado no documento a ser indexado.

Figura 8 – Fluxo simplificado do funcionamento e da coleta da Produção Técnico-científica digital no âmbito do Instituto

4 METODOLOGIA

Trata-se de um estudo longitudinal caracterizado como exploratório descritivo de abordagem quantitativa e do tipo documental.

Habitualmente, os estudos exploratórios observam os fatos desconhecidos e representam o início de uma pesquisa; em geral, estão associados aos modelos descritivos à medida que correlacionam os fatos observados (GIL, 1991).

Para selecionar o tipo de abordagem quantitativa ou qualitativa, consideraram-se os recursos materiais, temporais e pessoais disponíveis para realizar o trabalho, a questão tem implicações práticas, técnicas e empíricas (MINAYO e SANCHEZ, 1993). No presente estudo, justifica-se a opção pela abordagem quantitativa devido ao uso das técnicas bibliométricas.

A abordagem quantitativa vem da tradição das ciências naturais em que as variáveis a serem observadas são consideradas objetivas, reprodutíveis e generalizáveis, isto é, diferentes observadores obterão os mesmos resultados em observações distintas (HAYATI; KARAMI, SLEE, 2006). Citam-se algumas características (entre outras) da pesquisa quantitativa, como: a) obedece a um plano preestabelecido objetivando enumerar ou medir eventos; b) examina as relações entre as variáveis por meios experimentais, controlados com rigor; c) emprega habitualmente o instrumental estatístico para a análise dos dados e d) utiliza dados que representam uma população específica (amostra) a partir da qual os resultados são generalizados.

Na pesquisa documental examinaram-se materiais que ainda não receberam nenhum tratamento analítico *a priori*; os documentos são itens especialmente propícios para estudos do tipo longitudinal, estes se apoiam no tempo e descrevem uma série de características desenvolvidas durante certo período; têm importância significativa no contexto das ciências sociais (VALLEJO RUIZ, 2005).

O estudo de séries temporais é uma das técnicas mais utilizadas para descrever a “história” de determinada variável, como: a) obter as regularidades que se observam no comportamento de uma variável; b) prever seu comportamento

minimizando a incerteza, ou seja, pode ser definido como uma sucessão de observações correspondentes a uma variável em determinados momentos do tempo (VALLEJO RUIZ, 2005).

O *corpus* desta pesquisa é composto pelos dados extraídos da base de dados institucional do IPEN (PTC), entre outros itens de produção científica, como artigos de periódicos, de eventos, livros, capítulos de livros. Abrange as dissertações e teses produzidas e depositadas na unidade de informação do IPEN, no período de 1977 a 2009, constituindo-se na variável independente e quantitativa.

A preparação dos dados a serem utilizados no presente trabalho foi realizada anteriormente às análises propostas. Para tanto foi preparada uma base de trabalho especial, paralela, para facilitar a manipulação dos dados. Grande parte do tempo da elaboração do trabalho foi consumida no preparo, na complementação, na reformatação e na verificação da consistência dos dados.

Observou-se também que em um trabalho, que envolve dados bibliográficos para serem utilizados no processamento bibliométrico, a maior parte do tempo é consumida no preparo destes dados, corroborando a afirmação de CABENA (FIG.9), o qual realizou um estudo e afirma que 60% do tempo requerido para um trabalho destes é consumido no preparo e na formatação dos dados, no entanto, é uma condição prioritária para se obter sucesso em um estudo dessa natureza.

Figura 9 – Representação gráfica dos procedimentos envolvidos em um estudo utilizando os métodos bibliométricos para análise da informação.

Fonte: Adaptado de CABENA, P. et al., 1988, p.36

Utilizou-se o modelo de SPINAK (2008), (FIG.10) sobre as etapas que devem ser seguidas antes de se iniciarem os estudos métricos:

Figura 10 – Etapas a serem seguidas para se elaborar um estudo bibliométrico (SPINAK, 2008)

No caso deste trabalho as etapas de SPINAK (2008), foram adaptadas e se desenvolveram da seguinte forma:

- 1) definir o que se quer medir:** o grupo de dissertações e teses publicadas pelo Programa de Pós-graduação no IPEN;
- 2) definir o que se quer conseguir com a medida:** caracterizar o perfil científico das dissertações e teses;
- 3) determinar quais são as medidas disponíveis:** indicadores numéricos, temáticos e de produtividade;
- 4) criar combinações e correlações com essas medidas:** a) combinação das séries temporais numérica e temática; b) mensuração da produtividade por meio da correlação entre as palavras-chave da tese/artigo; c) mensuração do tempo transcorrido entre a defesa da tese e a publicação do artigo;
- 5) selecionar os instrumentos e processos para a coleta dos dados:** bases de dados, da biblioteca do IPEN, do INIS da IAEA, periódicos eletrônicos da CAPES e do Currículo Lattes; uso de softwares para tratamento dos dados;

- 6) validar a confiabilidade e a amplitude das medidas:** envio de questionários para os autores das teses solicitando a indicação do grau de correlação dos artigos.

4.1 Procedimentos adotados para construção da base de dados das dissertações e teses

Como fonte de dados foi utilizada a base de dados local PTC onde se encontram registrados todos os trabalhos técnico-científicos do IPEN, para tanto foram extraídos somente os dados das dissertações e teses e construída uma base à parte com o software Microsis, para facilitar o manuseio dos dados os procedimentos foram como segue:

1. seleção/extração dos documentos: dissertações e teses (1670) do período de 1977 a 2009, registradas na base de dados da biblioteca do IPEN (PTC);
2. construção de uma base específica para reformatação e preparação dos dados;
3. verificação da consistência e da estrutura dos dados, correção de nomes de autores e orientadores, erros de digitação, padronização dos nomes da Instituição, completeza dos dados, eliminação da duplicidade de palavras-chave;
4. tratamento informatizado dos dados e preparação para análise bibliométrica por meio de softwares especializados. O *Dataview*²¹ e o *Infotrans*²² foram utilizados para a produção de listas de frequência e correção e reorganização dos dados;
5. adoção de metacategorias para a distribuição temática das dissertações e teses utilizando a tabela *Subject Categories and Scope Descriptions*

²¹ Software elaborado pelo Centre de Recherche de Marseille na França, versão DOS, produz listas de frequência e matrizes, permitindo checar a consistência dos dados.

²² Software alemão da empresa Information und Kommunikation (luk), versão DOS, permite a padronização automática de campos, complementação de campos com dados externos e reorganização geral da base de dados para a eliminação de registros duplicados.

do INIS (INTERNATIONAL ATOMIC ENERGY AGENCY, 2002). Esta tabela é composta de 45 categorias temáticas, no caso dos trabalhos do IPEN estes se distribuem em 33 categorias, isto é, a tabela do INIS é mais ampla e nem todas as categorias se aplicam; a tabela é utilizada por todos os indexadores dos sistemas que contribuem com a base INIS no mundo. Fornece orientação sobre a abrangência temática geral da categoria com explicações detalhadas sobre os subtemas que se aplicam ou não dentro de cada grande tema; esta orientação garante que a base mantenha o foco no tema nuclear e ciências relacionadas. É atualizada periodicamente pelo quadro de especialistas do INIS;

6. construção de matrizes²³ temáticas e temporais, com os dados coletados, utilizando o programa Excel para tabulação e emissão de gráficos dos dados.

4.2 A produtividade das teses com relação à publicação de artigos

É importante salientar que ao se abordar este aspecto de produtividade das teses, ressalta-se a publicação de artigos produzidos a partir da elaboração do trabalho, visto que o conceito de produtividade deste tipo de literatura é bem mais amplo, o qual inclui as aplicações práticas, como por exemplo, a produção de patentes e o desenvolvimento de novas tecnologias que possam advir da elaboração de um trabalho acadêmico como é o caso das teses.

MORAVCSIK (1988), comenta que há três conceitos em ciência que devem ser entendidos antes de realizar um estudo:

- ✓ a *atividade* é o esforço, a energia despendida, mas sem haver referência em que sentido se avança na tarefa que se quer executar, definida também como o consumo relevante dos insumos (*input*);

²³ Para efeito deste trabalho, utiliza-se o termo “matriz” para nomear a tabela em que se reuniram todos os dados necessários para tabulação e emissão das diversas listas e relatórios.

- ✓ a *produtividade* significa o grau em que este consumo produz resultados relevantes (*output*);
- ✓ e o *progresso*, que mede o grau em que essa produtividade conduz mais rapidamente para se atingir os objetivos específicos estabelecidos.

MORAVCSIK (1988), também comenta que produzir indicadores para medir a atividade é mais simples, não há necessidade de ser especialista em ciência, no entanto medir produtividade e principalmente progresso é mais complexo e requer maior compreensão da atividade científica.

O que se deseja evidenciar é que o processo de produção de conhecimento (baseado em pesquisas) necessita utilizar os canais de comunicação científica para a divulgarem os resultados obtidos e assim submetê-los à avaliação dos seus pares. Assim a produção científica é ampla e atualmente utiliza diversas mídias. Abrange comunicações, apresentações, pôsteres, palestras, relatórios, artigos apresentados em eventos ou publicados em revistas e outras formas de comunicação..

A publicação de artigos é para maioria dos pesquisadores não só um indicador, mas um produto final de todo um trabalho específico, um resultado que somente é reconhecido se for comunicado. O ciclo do conhecimento só se completa quando este resultado é avaliado e aceito por outros pesquisadores da mesma área (BRAGA, 1974).

De acordo com MEADOWS (1999), em termos de comunicação científica as duas características mais importantes do pesquisador são a quantidade e a qualidade de informações que comunica. E cita como uma das medidas de quantidade (produtividade) o número de artigos publicados em periódicos. Uma forma de avaliar a qualidade seria o nível de interesse dos outros pesquisadores pela pesquisa publicada (índice de citação).

Entretanto, reconhece que ambos são aspectos complexos de serem abordados quando se trata de mensuração; no caso específico da produtividade, esta exige ponderações adotadas para as diferentes categorias de publicações. Isto se deve ao fato de as áreas de conhecimento adotarem preferencialmente diferentes canais de comunicação científica, como por exemplo, a área de humanas prefere publicar o

resultado das suas pesquisas em forma de livros, e área de medicina ou exatas prefere artigos de periódicos, uma das questões seria, por exemplo, em termos brutos quantos artigos correspondem a um livro publicado? Há estudos empíricos, porém a dificuldade de como fazer essa comparação entre a produtividade das diferentes áreas permanece, e de acordo com MEADOWS (1999), ainda não foi resolvida, embora já tenham sido realizadas várias tentativas.

Note-se que, neste trabalho, para se obterem os dados sobre a produtividade foram utilizados somente os artigos de periódicos declarados pelos autores das teses e, no caso todas as que foram defendidas e homologadas no Programa de Pós-graduação do IPEN, no período 1977-2009, por se entender que a tese encerra um ciclo de pesquisa em que a produção de artigos é inerente a esse processo.

4.3 Procedimentos adotados para a construção das tabelas matrizes dos artigos

Objetivando identificar o número de artigos publicados em periódicos nacionais e internacionais pelos autores das teses, foi realizado um levantamento no Currículo Lattes (CL) durante 12 meses e foi finalizado em outubro de 2010; a etapa de coleta foi a que menos consumiu tempo, em comparação às outras etapas, como, por exemplo, a da consistência, localização, complementação dos dados, estas foram intercaladas com a etapa de coleta. Descrevem-se as etapas:

- ✓ seleção das teses que se constituíram no *corpus* desta fase do trabalho; período de 1977 até 2009, no qual foram defendidas 536 teses no Programa de Pós-graduação do IPEN; desse total, alguns autores não tinham Currículo Lattes (CL) ou não tinham artigos no período estabelecido (5 anos antes e 5 anos depois), ou ainda não tinham nenhum artigo publicado; assim, o universo de teses verificado foi reduzido para 401 itens;
- ✓ definição de um período de 5 anos antes e 5 anos depois da apresentação da tese como ponto de partida para levantamento dos artigos publicados. Esse período foi selecionado como um período médio adequado, uma vez que

houve modificações ao longo da história do Programa e as teses já tiveram diferentes prazos de tempo estipulados pela PG para sua elaboração (1977 – prazo 8 anos; 1997 – prazo 5 anos; atualmente – prazo 4 anos);

- ✓ verificação de cada um dos autores de teses no Currículo Lattes, no item “Artigos completos publicados em periódicos”. Em um primeiro momento Identificaram-se 2.556 artigos; dos quais foram excluídos aqueles que não se configuravam em artigos de periódicos (artigos de eventos, publicações IEA, IPEN, e outros). Excluíram-se, também, do total de artigos, 345 itens que não foram localizados (dados incompletos, não conferiam, revista não identificada ou não localizada), restando finalmente 2.211 artigos nacionais e internacionais para serem analisados neste trabalho. Foram identificados e coletados também, dentro deste universo, 390 artigos (17%), onde não constava o nome do IPEN no campo filiação; e de acordo com a política de conformação da base, esses artigos não foram incorporados à base PTC²⁴, contudo, foram contabilizados e indexados para fins deste trabalho, de acordo com a declaração do autor no Currículo Lattes (CL);
- ✓ construção de uma matriz em tabela Excel, com as teses e seus respectivos artigos todos codificados com uma numeração única, IPEN-Doc (utilizado para controle interno das publicações na base PTC), tanto as teses como os artigos dispõem de dados completos como autor, orientador, ano de publicação, título, volume, palavras-chave, código de identificação e categoria de assunto (Apêndice B); optou-se por utilizar a planilha do software Excel, em vez de construir uma base de dados especial com estes dados, para facilitar o trabalho de tabulação e construção de gráficos e das tabelas;
- ✓ padronização dos dados, ação similar àquela efetuada quando da preparação da base especial para as dissertações e teses, exemplo: sobrenome de autores, orientadores, dados dos artigos, títulos de periódicos;

²⁴ A base PTC conta atualmente com 3.173 artigos nacionais e internacionais. Muitos dos artigos coletados no Currículo Lattes já se encontravam depositados na base, aproximadamente 280 (12%) artigos foram recuperados e acrescentados à base.

- ✓ identificação e localização e incorporação dos artigos quando não se encontravam depositados na base PTC;
- ✓ correção das palavras-chave (PCs)²⁵. A este item foi dedicada atenção especial, uma vez que estas serviriam como ponto de correlação. Assim, foi verificada a consistência de 3.332 palavras-chave, com relação a: a) duplicidade; b) grafia; c) hifenização; d) número (singular, plural) e e) separadores sempre utilizando o Tesouro INIS como parâmetro e utilizando o software *Dataview* para emissão das listas;
- ✓ adoção de procedimentos para fins de padronização e análise deste trabalho, por exemplo: para os autores com sobrenomes espanhóis, sobrenomes compostos ou que adotaram várias formas de assinatura em seus trabalhos, foi estabelecida uma só forma de entrada, sempre de acordo com o que estava registrado na tese; para os casos de sobrenomes de autoras que deixaram de adotar ou incorporaram outro sobrenome após mudança de estado civil, estabeleceu-se para fins deste trabalho que se manteria o mesmo sobrenome da época da tese;
- ✓ utilização dos softwares *C/Isis*²⁶ para tratamento e manipulação dos dados, o *Excel* para cálculo e produção de tabelas e o *Dataview* para emissão de listas e “higienização” dos dados.

Na segunda etapa os procedimentos adotados para a definição dos critérios para estabelecer a correlação foram:

- ✓ seleção de uma amostra de 128 autores de teses (32%) objetivando testar o método automático²⁷ e validar os resultados com a opinião dos autores; foi enviado um questionário no formato fechado autoexplicativo e autoaplicável (Apêndice A);

²⁵ Na base PTC, devido ao uso do Tesouro do INIS, utilizam-se as palavras-chave sempre em inglês, independentemente do idioma do documento a ser indexado.

²⁶ Biblioteca de funções desenvolvida pela BIREME em linguagem C para permitir a operacionalização de bases de dados ISIS sem a necessidade de instalação ou uso do software como o CDC/ISIS – Microisis/Winisis (UNESCO): Disponível em <http://bvsmodelo.bvshalud.org/php/level.php?lang=pt&component=28>

²⁷ Para fins deste trabalho, denominou-se “método automático” a correlação estabelecida pelas palavras-chave entre a tese e o artigo, pela combinação binomial.

- ✓ a seleção dos autores foi aleatória, porém procurando distribuir a seleção por toda a tabela matriz, posteriormente alguns autores selecionados tiveram que ser substituídos, devido à impossibilidade de obter o endereço do correio eletrônico, este fato aconteceu com os autores de teses mais antigas, neste caso a opção foi selecionar o autor seguinte na tabela matriz. O questionário foi elaborado de forma simples e direta com todos os dados das teses e dos artigos coletados no CL (Apêndice A); nele foi solicitado que o autor indicasse, entre quatro alternativas, o nível de correlação entre a pesquisa desenvolvida; na tese e no artigo publicado sendo 1 – correlação mais forte e 4 – sem nenhuma correlação. Antes de enviar o questionário foi realizado um pré- teste com doze autores das teses, o resultado do pré-teste mostrou que um conjunto de 4 alternativas (em vez de 2 ou 3) para as respostas representava melhor a opinião dos autores.
- ✓ envio dos questionários aos autores via correio eletrônico; dos 128 autores contatados (30% do total de autores), 100 responderam, até o momento de fechar a tabulação (posteriormente foram recebidos mais 4 formulários), apontando o índice de correlação de 397 artigos;
- ✓ tabulação dos resultados das respostas dos autores das teses e verificação do índice de correlação assinalado pelos autores; definição dos critérios para estabelecer a correlação pelo método automático¹⁸ de coocorrência;
- ✓ aplicação da metodologia a todo o *corpus* de estudo, isto é, foi definido o critério de adotar três ou mais palavras-chave como índice de correlação, aplicou-se esta metodologia para todos os artigos coletados no CL, no caso 2211.

4.4 A técnica de coocorrência

Neste trabalho, será utilizada a coocorrência de palavras-chave do campo indexação dos documentos, para identificar a correlação dos artigos com as teses, com base no fato de que na área nuclear, é utilizado um único instrumento de

indexação, um vocabulário controlado (Tesauros) para todos os documentos que compõem a base de dados da área nuclear o INIS,

É oportuno comentar que neste trabalho se adotou o termo palavras-chave ao invés de descritores, porque tradicionalmente desde que foi criado em 1974, o sistema de informação utilizado no IPEN adota este termo, porém sempre se referindo as palavras-chave extraídas de um vocabulário controlado, no caso o Tesouro do INIS.

A técnica de coocorrência de palavras foi utilizada como instrumento para a identificação de uma correlação mais próxima entre a tese e os artigos publicados antes, durante ou após a elaboração da tese. A correlação não foi utilizada para confeccionar *clusters* ou redes sociais, destinadas a mapear a estrutura ou dinâmica da ciência.

Nessa técnica, são identificadas as palavras em determinado contexto, sem se importar com o conteúdo semântico, e sim com a frequência das combinações dos pares COURTIAL (1994); HE (1999); DING et al. (2001); ROKAYA et al. (2008); NEFF e CORLEY (2009). Em outras palavras, no momento do uso da coocorrência de termos, não se considera o significado ou a definição das palavras, mas o fato de as palavras estarem presentes *em pares* nos documentos em que se pretende dimensionar a correlação.

A combinação das palavras, duas a duas, possibilita a identificação de palavras coocorrentes de um número finito de elementos do conjunto finito de palavras-chave do documento. Por exemplo: em um documento onde aparecem as palavras A, B, C e D, os diferentes pares resultantes serão AB, AC, AD, BC, BD e CD (total de 6 pares coocorrentes). O número binomial serve como fórmula geral para determinação da quantidade de combinações de pares de palavras no documento:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Como na técnica de coocorrência se estabelecem **pares** de palavras, tem-se:

$$\binom{n}{2} = \frac{n!}{2!(n-2)!}$$

Sendo ***n*** o total de palavras, resultando no total de possíveis combinações. Para reforçar o exemplo anterior, com as quatro palavras A, B, C e D, tem-se:

$$\binom{4}{2} = \frac{4!}{2!(4-2)!} = \frac{4.3}{2} = 6$$

A título de exemplo na FIG. 11 a seguir, podem-se observar a aplicação da fórmula e o número resultante de coocorrências:

Figura 11 – Exemplo de correlação entre o número de palavras-chave (PCs) e o número de coocorrências

Nº Palavras-chave (PCs)	Nº de coocorrências
2	1
3	3
4	6
5	10
6	15
7	21
8	28
9	36
10	45
11	55
12	66
13	78
14	91
15	105
16	120
17	136
18	153
19	171
20	190

Apresenta-se também um exemplo simplificado (FIG. 12) de como o número binomial foi utilizado para trabalhar com as palavras-chave, no caso trata-se de uma tese do autor que publicou esse artigo, coletado do Currículo Lattes, no período estabelecido para estudo, cinco anos antes e cinco anos depois da defesa da tese.

Neste exemplo observa-se que, pela indexação, há seis palavras-chave coincidentes entre a tese e o artigo publicado:

TÍTULO /TESE	PALAVRAS-CHAVE/TESE	TÍTULO/ARTG	PALAVRAS-CHAVE/ARTIGO
AVALIAÇÃO DO CONTEÚDO DE ALGUNS ELEMENTOS ESSENCIAIS E TÓXICOS EM DIETAS DE CRIANÇAS E IDOSOS PELO MÉTODO DE ANÁLISE POR ATIVAÇÃO COM NEUTRONS	NEUTRON ACTIVATION ANALYSIS; DIET; TOXIC ELEMENTS; CHILDREN; AGED ADULTS; MULTI-ELEMENT ANALYSIS; RADIOCHEMICAL SEPARATION; NUTRIENTS; TRACE AMOUNTS	DETERMINATION OF VARIOUS NUTRIENTS AND TOXIC ELEMENTS IN DIFFERENT BRAZILIAN REGIONAL DIETS BY NEUTRON ACTIVATION ANALYSIS	DIET; FOOD; BRAZIL; NEUTRON ACTIVATION ANALYSIS; MULTI-ELEMENT ANALYSIS; NUTRIENTS; SEMIMETALS; TRANSITION ELEMENTS; TRACE AMOUNTS; TOXIC ELEMENTS

Figura 12 – Exemplo de coocorrência de palavras-chave entre uma tese e um artigo

- 1- neutron activation analysis (A) 2- trace amounts(B)
 3- multielement analysis(C) 4-toxic elements (D)
 5- diet (E) 6- nutrients (F)

Total 6 PCs

A,B,C D,E,F = AB, AC, AD, AE, AF

BC,BD,BE,BF

CD, CE, CF

DE, DF

EF

Aplicando-se a fórmula do binomial para 6 de palavras-chave, teremos:

$$\binom{6}{2} = \frac{6!}{2!4!} = 15 \text{ testes}$$

Isto é, para 6 palavras-chave coincidentes, consideram-se 15 coocorrências.

4.5 Tempo para publicação e a visibilidade dos artigos

Outros dois aspectos foram verificados nesse material: a) o tempo transcorrido entre a defesa da tese e a publicação do artigo a fim de comprovar a percepção pessoal da autora do trabalho de que esse tempo estaria sendo abreviado, senão antecipado, fato positivo para o processo de comunicação científica; e b) a indexação dos títulos de periódicos, na base de dados internacional Web of Science (WoS), utilizados para publicação dos artigos, isto é, a visibilidade internacional dos canais de comunicação científica.

Com relação ao item b, esclarece-se que, inicialmente, no planejamento desta pesquisa a intenção era verificar e acrescentar o Fator de Impacto (FI) de cada título. Isto não foi viável, porque o *Journal Citation Report* (JCR)²⁸ disponibiliza estes dados somente a partir de 2001, e para se ter um indicativo mais homogêneo e real, deveriam ter sido utilizados os valores do FI correspondentes ao ano da publicação do artigo, presumindo-se que este poderia variar de valor, dependendo do ano de publicação do artigo.

A fim de se ter uma visão mais verdadeira sobre este aspecto foi elaborado um quadro demonstrativo com os valores do FI desde 2001, data em que os dados passaram a estar disponíveis no JCR.

Para tanto, selecionaram-se os dez títulos internacionais com maior número de artigos dentro do período estudado, e se verificou o valor do FI anual (TAB. 1).

Conclui-se que nesses títulos pesquisados, há alterações nos valores do FI dos títulos citados, ao longo dos anos; esta alteração acontece de forma intermitente com pequena variação, porém suficiente para interferir e causar vieses nos resultados. No entanto, o fato de os títulos de periódicos estarem indexados na Web of Science (WoS) já indica que os canais utilizados para comunicação das pesquisas têm visibilidade internacional, assim se julgou conveniente acrescentar este dado para completar a análise dos títulos.

²⁸ *Journal Citation Report*, publicado pelo ISI da Thompson, lista os títulos com seu respectivo Fator de Impacto; foi utilizado praticamente este instrumento institucionalizado para se estabelecer o ranking dos periódicos internacionais, além de ser bem aceito pela comunidade científica.

Tabela 1 – Demonstrativo da evolução do Fator de Impacto dos 10 títulos de periódicos que mais produziram artigos no período de 2001 a 2009

Periódico	nº artigos	2001	2002	2003	2004	2005	2006	2007	2008	2009
Materials Science Forum	126	0.461	0.613	0.602	0.498	0,399	-	-	-	-
Radiat. Physics and Chemistry	65	0,819	0,738	0,693	0,890	0,725	0,868	0,934	0,882	1,149
J. of Radioanal. and Nucl. Chemistry	51	0.398	0.502	0.472	0.457	0.460	0.509	0,499	0,659	0,631
Applied Radiation and Isotopes	38	0,635	0,768	0,690	1,000	0,757	0,924	1,034	1,114	1,094
Brazilian Journal of Physics	33	0.719	0,678	0,277	0,435	0,445	0,494	0,478	0,512	0,575
Key Engineering Materials	32	-	0,497	0,284	0,278	0,224	-	-	-	-
J. of Alloys and Compounds	31	0.953	1,014	1,080	1,562	1,370	1,250	1,455	1,510	2,135
Nucl. Instrum. Meth. Phys. Res. B	31	1.041	1.158	1.041	0,997	1.181	0.946	0.997	0.999	1.156
Nucl. Instrum. Meth. Phys. Res. A	26	1.026	1.167	1.166	1.349	1.224	1.185	1.114	1.019	1.317
J. of Applied Physics	25	2.128	2.281	2.171	2.255	2.498	2.316	2.171	2.201	2.072

Fonte: Tabela construída a partir dos dados do JCR, ISI Thompson Reuters 2009, consultado em 2011.

5 RESULTADOS

Para uma melhor visualização, dividiu-se a apresentação dos resultados em duas etapas. Na primeira etapa serão apresentados os resultados numéricos e temáticos obtidos a partir das diferentes análises de tendências realizadas com **as dissertações e teses** produzidas pelo programa de PG do IPEN. Esclarece-se que os resultados numéricos (FIG. 13) dizem respeito às dissertações e teses publicadas e depositadas na base de dados Produção Técnico-Científica (PTC) da biblioteca do IPEN. É possível haver pequena discrepância com relação aos totais das dissertações e teses, disponibilizados em gráficos no site da Pós-graduação do IPEN, isto se deve principalmente à data de homologação ou indisponibilidade do trabalho na biblioteca, na época de fechamento da coleta dos dados. A segunda etapa de apresentação dos resultados será reservada para mostrar os resultados da verificação dos artigos produzidos pelas teses.

5.1 Distribuição temporal e temática das dissertações e teses

Figura 13 – Distribuição temporal das dissertações e teses defendidas no Programa de Pós-Graduação do IPEN no período de 1977 a 2009

Tabela 2 – Distribuição das Dissertações e Teses produzidas pelo Programa de Pós-Graduação do IPEN no período de 1977 a 2009 por categorias temáticas e percentual de participação

Código da categoria no INIS	Título das categorias	Nº de diss. e teses	%
S36	Ciência dos Materiais	335	20,06
S37	Química Orgânica e Inorgânica	186	11,14
S21/22	Reatores	158	9,46
S63	Aplicações e Efeitos da Radiações	129	7,72
S38	Radioquímica	96	5,75
S61	Proteção Radiológica	95	5,69
S46	Instrumentação Nuclear	87	5,21
S60	Aplicações em Ciências da Vida	83	4,97
S54	Meio Ambiente	80	4,79
S62	Medicina Nuclear	73	4,37
S75	Física Condensada	70	4,19
S07	Radioisótopos	48	2,87
S99	Gestão	47	2,81
S73	Física Nuclear	43	2,57
S11	Ciclo do Combustível	43	2,57
S12	Rejeitos	34	2,04
S42	Engenharia	16	0,96
S43	Aceleradores	9	0,54
S58	Geociências	7	0,42
S30	Conversão de Energia	4	0,24
S29	Energia – planejamento, política	4	0,24
S09	Biomassa	4	0,24
S08	Hidrogênio	4	0,24
S71	Mecânica Quântica	3	0,18
S74	Física Atômica e Molecular	2	0,12
S72	Física de Partículas	2	0,12
S70	Física do Plasma	2	0,12
S98	Salvaguardas/ prot. Física	1	0,06
S47	Instrumentação não nuclear	1	0,06
S32	Energia – conservação	1	0,06
S25	Energia – armazenamento	1	0,06
S10	Combustíveis Sintéticos	1	0,06
S03	Gás Natural	1	0,06
Total		1670	100,00

Na TAB. 2 se apresenta a distribuição das dissertações e teses por categoria temática e numérica em ordem decrescente, bem como o percentual de contribuição de cada categoria. O primeiro código à esquerda da tabela (S36) se refere a uma codificação própria do INIS. Foi a partir desta tabela que se efetuaram as análises de trajetória e de tendências.

Foram adotadas essas metacategorias temáticas utilizando-se a tabela do sistema internacional INIS, o *Subject Categories and Scope Descriptions* (INTERNATIONAL ATOMIC ENERGY AGENCY, 2002), para agrupar e categorizar tematicamente as dissertações e teses durante o período de 1977 a 2009. Nos casos em que os trabalhos apresentavam duas categorias temáticas (3%), a opção foi sempre pelo primeiro tema, de acordo com a orientação dos manuais de categorização do INIS.

Figura 14 – Distribuição numérica e temática das dissertações e teses com as categorias nucleares agrupadas, período 1977-2009

Na FIG. 14 se apresenta a distribuição temática das dissertações e teses produzidas no período de 1977 a 2009. Em um primeiro cenário os temas nucleares estão agrupados. É oportuno comentar que, devido à opção de utilizar a tabela temática estabelecida pelo INIS, e por se tratar especificamente da área em questão, a tabela se desdobra nas especificidades temáticas da área nuclear. Dessa forma, os temas que tem grande envolvimento com as técnicas nucleares foram agrupados para fins de melhor visualização comparativa com os temas que foram mantidos em plano mais geral. Assim, dentro da denominação nuclear estão agrupados 11 temas

cujos trabalhos tratam de: proteção radiológica; aplicações e efeitos das radiações; instrumentação nuclear; reatores nucleares; medicina nuclear; radioquímica; salvaguardas; ciclo do combustível; rejeitos; radioisótopos; aplicação das técnicas nucleares nas ciências da vida, e os vários subdomínios da área de Física (física nuclear, matéria condensada, física geral, física do plasma, quântica); os temas sobre energia (planejamento, conservação, armazenamento) também foram agrupados, por se constituírem temáticas muito semelhantes e em algum dos casos, apresentando um número ainda pequeno de trabalhos. Com relação a essa decisão, julgou-se conveniente manter e apresentar os dois cenários (com agrupamento e sem agrupamento), uma vez que a estratificação temática desagrupada facilitou a análise de trajetória e formação de *clusters* (TAB. 3), como também favoreceu a correlação com os fatos históricos ou políticos tornando-os mais elucidativos. Ainda neste cenário, agrupando as categorias que compõem a área nuclear se apresenta a mesma distribuição em percentuais (FIG.15), observando-se que as dissertações e teses elaboradas nos temas nucleares se constituem na maioria (51%), seguida pela forte presença do tema ciência dos materiais (20%). Enquanto as outras categorias apresentam uma distribuição percentual mais homogênea.

Figura 15 – Distribuição percentual das dissertações e teses, mantendo-se as categorias temáticas nucleares agrupadas, período 1977-2009

Com o objetivo de obter uma visualização mais específica da trajetória das categorias temáticas, apresentam-se também os dez temas com maior número de

dissertações e teses no período 1977-2009, desta vez sem o agrupamento dos temas nucleares, no entanto, as duas categorias reatores foram agrupadas por se tratarem de temas muito relacionados, diferindo apenas na abordagem; bem como os cinco subdomínios da física, como física nuclear, física atômica, física do estado sólido, física do plasma e física quântica, também foram mantidos agrupados por apresentarem um número pequeno de trabalhos e assim evitar a dispersão.

Figura 16 – Distribuição temática, percentual das 10 categorias com maior número de dissertações e teses, sem agrupamento, período 1977-2009

Na FIG. 16, apresenta-se o percentual dessa distribuição. Note-se que há uma distribuição mais homogênea com predominância da categoria de ciência dos materiais.

Com os seguintes gráficos pretende-se demonstrar como os trabalhos se distribuíram tematicamente no início do Programa de PG, período 1977-1984 (este primeiro período foi agrupado em 8 anos, para favorecer a formação dos quinquênios nos anos subsequentes) (FIG. 17).

Observa-se que inicialmente os trabalhos se distribuíam em 17 categorias temáticas com grande predominância do tema química; os motivos são analisados na parte de Discussão (p.139) deste trabalho. Já no quinquênio seguinte (1985-1989), um período politicamente instável do Programa de pós-graduação (ver comentários na parte de Discussão, p. 139 deste trabalho), não só o número de trabalhos decresce (FIG.13) como também o número de categorias temáticas (15), onde os trabalhos se distribuem (FIG.18).

A partir da década de 1990 há uma expansão temática, a distribuição se torna mais homogênea justificando a multidisciplinaridade da área nuclear e evidenciando as ciências relacionadas, os trabalhos se distribuem em 21 categorias temáticas (FIG. 19), porém ainda sem perder o foco no tema nuclear, mantendo esta abrangência até os dias atuais (2005-2009).

Demonstra-se assim como aconteceu a “dinâmica temática”, dentro do Programa da Pós-graduação, ao longo do período estudado, 1977-2009; por exemplo, observe-se a movimentação das categorias temáticas áreas de Química, Ciência dos Materiais, Reatores, Física e Efeitos das Radiações, ora em evidência, ora com desempenho desacelerado, influenciadas pelos momentos políticos ocorridos na área maior (energia nuclear) no país. Também se percebeu o aparecimento de novas categorias temáticas, como, por exemplo, foram desenvolvidos trabalhos nos temas de biomassa, hidrogênio, aceleradores e energia.

Figura 17 – Distribuição temática das dissertações e teses em 17 categorias no período inicial do programa de PG, período **1977-1984**

Figura 18 – Distribuição temática das dissertações e teses em 15 categorias no período **1985-1989**

Figura 19 – Distribuição temática das dissertações e teses em 21 áreas, no último período estudado **2005-2009**

5.2 Trajetória e tendências temáticas

Com o objetivo de se conhecer mais detalhadamente o comportamento temático das 1.670 dissertações e teses foram calculadas medidas descritivas como a mediana e o desvio padrão para cada uma das 33 categorias temáticas.

A partir dessa primeira análise observou-se um comportamento anual errático, o qual motivou a agregação desses dados em períodos quinquenais, aproximadamente; somente o período inicial de 1977 a 1984 foi mantido agrupado

com 8 anos, com o objetivo de favorecer a formação de quinquênios nos últimos anos e também devido ao número de trabalhos neste período ser menor. Posteriormente, foi utilizada uma técnica multivariada denominada *análise de agrupamento*, ou *análise de clusters*. Esta técnica permitiu identificar a quantidade de trabalhos de cada quinquênio das áreas e sinalizou as séries com menores distâncias ente si. Adicionalmente, visando avaliar a consistência dos resultados, foi usada outra técnica estatística denominada *análise de trajetória*. Essa técnica também tem por objetivo identificar grupos de sequência e com evolução similar ao longo do tempo, porém associando-os (baseia-se no modelo de regressão polinomial de JONES et al., 2001).

Sintetizando, as duas são técnicas estatísticas de análise multivariada e neste trabalho foram utilizadas para estabelecer padrões de comportamento das diferentes áreas temáticas com base em sua trajetória passada.

De forma geral, ambas apresentaram resultados bastante similares com diferença um pouco mais acentuada em algumas séries, o que determinou a utilização das duas técnicas para validar os resultados.

Para a análise estatística foram utilizados os softwares SPSS v.16.0 e SAS 9.1; para construção de matrizes e obtenção dos gráficos, utilizou-se o Excel 2007.

Na TAB. 3, visualizam-se e descrevem-se as características dos grupos formados:

O grupo 1, formado pela categoria temática de ciência dos materiais, apresenta um nível altíssimo de produção de dissertações e teses no período estudado e com tendência a forte crescimento.

O grupo 2, composto por duas categorias temáticas (química e reatores), apresentou um comportamento peculiar em formato de “U”, ou seja, teve um nível médio no primeiro quinquênio, passando posteriormente por uma queda acentuada de produção no segundo, recuperando-se e apresentando uma tendência alta e crescente.

De forma geral, o grupo **3**, composto por cinco categorias (radioquímica, meio ambiente, aplicações da energia nuclear nas ciências da vida, efeitos das radiações e proteção radiológica), apresenta um nível médio de produção um pouco mais elevado do que os dois últimos grupos, porém, com forte tendência de elevação, a partir do terceiro quinquênio.

O grupo **4**, composto por oito categorias (fontes de radiação; ciclo do combustível; rejeitos; instrumentação; medicina nuclear; física nuclear e física da matéria condensada, e gestão), apresenta um desempenho médio e uma tendência de crescimento, porém, moderada.

Finalmente o grupo **5**, composto pelas áreas de: gás natural; hidrogênio; biocombustíveis; combustíveis sintéticos; geociências; armazenamento, planejamento, conversão e conservação de energia; aceleradores; instrumentação não nuclear e física do plasma, mecânica quântica; física atômica; salvaguardas, apresenta séries com uma produção baixa e errática.

Tabela 3 – Resultado do agrupamento temático das dissertações e teses, após a análise de trajetória e análise de clusters

Grupo	Categoria temática	Característica
G1	Ciência dos materiais	Altíssima produção com tendência de elevação
G2	Química e Reatores	Produção alta, comportamento em U, tendência de elevação
G3	Radioquímica, Meio ambiente, Aplicações em ciências da vida, Efeitos da radiação e Radioproteção	Produção média com tendência de forte elevação
G4	Radioisótopos e Fontes de radiação, Ciclo do comb., Rejeitos, instrumentação; Med. nuclear; Física nuclear e matéria condensada, Gestão	Produção média com tendência de elevação moderada
G5	Outras categorias	Produção baixa e errática

Figura 20 – Evolução e tendências dos 5 grupos temáticos formados pelas dissertações e teses, após a aplicação da técnica de análise de clusters, distribuídas pelos quinquênios estudados 1977-2009

Na FIG. 20, apresenta-se de forma gráfica a evolução e as tendências dos cinco grupos temáticos formados pelas dissertações e teses, após a aplicação da técnica de análise de clusters, distribuídas nos seis quinquênios estudados; o primeiro agrupamento foi mantido maior (8 anos) objetivando facilitar a distribuição quinquenal dos períodos subsequentes; é possível observar também, mais claramente, as tendências dos grupos.

Figura 21 – Visualização gráfica do nível de produção das séries temáticas distribuídas em quinquênios

Na representação visual (FIG. 21), observa-se o comportamento dos grupos em nível de produção de trabalhos. O grupo 3, apesar de apresentar um nível médio similar ao do grupo 4 nos três primeiros quinquênios (Q1, Q2, Q3), reage a partir do terceiro quinquênio passando a apresentar uma elevação, mudando para um nível similar ao do grupo 2, passando a acompanhar este grupo nos dois últimos quinquênios (Q5, Q6).

Na TAB. 4, apresentam-se os resultados da análise de cluster hierárquica utilizando o método de Ward. É possível observar a formação de cinco grupos de dissertações e teses, apresentados em quinquênios. A identificação dos padrões se baseia mais na comparação do que a partir da fixação de um valor limite fixo para cada uma das classes. Denominou-se altíssima a categoria que apresentava um nível de produção maior do que todas as demais, sendo a denominação alta dada às categorias seguintes que apresentaram produção similar estatisticamente e inferior à categoria denominada altíssima.

Assim, na tabela (TAB. 4) onde se vê:

- ✓ *N* – corresponde ao número de categorias envolvidas em cada grupo. Grupo 1 – Ciência dos Materiais; Grupo 2 – Química e Reatores; Grupo 3 – radioquímica, meio ambiente, aplicações da energia nuclear nas ciências da vida, efeitos das radiações e proteção radiológica; Grupo 4 – fontes de radiação; ciclo do combustível; rejeitos; instrumentação; medicina nuclear; física nuclear e física da matéria condensada, gestão; e o Grupo 5 composto por mais 15 categorias).
- ✓ *Média* – o número médio de trabalhos das categorias que participam dos grupos, apresentadas por quinquênio.
- ✓ *Desvio padrão e erro padrão de média* para cada grupo.

Tabela 4 – Resultados da análise hierárquica de Ward, das categorias temáticas das dissertações e teses, período 1977-2009 com classificação de níveis

Cluster com variáveis originais Ward		DISSERTAÇÕES E TESES						
		1977-1984	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009	
1	N	1	1	1	1	1	1	
	média	21,00	15,00	35,00	56,00	92,00	116,00	Nível muito alto
	Desvio padrão	.	.	.	.	.	.	
	Erro padrão da média	.	.	.	.	.	.	
2	N	2	2	2	2	2	2	
	média	43,50	9,00	18,50	26,00	27,50	47,50	Nível alto
	Desvio padrão	2,121	2,828	,707	5,657	3,536	21,920	
	Erro padrão da média	1,500	2,000	,500	4,000	2,500	15,500	
3	N	5	5	5	5	5	5	
	média	6,80	3,80	8,20	13,80	24,40	39,60	Nível alto
	Desvio padrão	6,099	3,421	2,588	6,535	9,737	5,683	
	Erro padrão da média	2,728	1,530	1,158	2,922	4,354	2,542	
4	N	8	8	8	8	8	8	
	média	7,38	4,38	5,88	7,38	12,75	18,00	Nível médio
	Desvio padrão	4,033	3,623	2,997	4,033	6,042	6,908	
	Erro padrão da média	1,426	1,281	1,060	1,426	2,136	2,442	
5	N	17	17	17	17	17	17	
	média	,41	,29	,18	,65	1,00	1,12	Nível baixo
	Desvio padrão	1,228	,849	,529	1,272	1,118	1,269	
	Erro padrão da média	,298	,206	,128	,308	,271	,308	
Total	N	33	33	33	33	33	33	
	média	6,30	2,79	4,94	7,48	11,76	17,33	
	Desvio padrão	11,004	4,014	7,525	11,694	18,007	24,671	
	Erro padrão da média	1,916	,699	1,310	2,036	3,135	4,295	

As análises fazem referência a fatos já ocorridos com as séries temporais, como comportamento ao longo do tempo, crescente, decrescente, estável. A abordagem de tendências fornece a ideia de perspectiva futura, contudo são sempre

tendências. Para se prever quantidades específicas, seriam necessárias análises estatísticas com técnicas mais complexas.

Estudos temporais dessa natureza contribuem para se conhecer a trajetória percorrida e a tendência de crescimento dos grupos, porém, o mais importante é identificar o comportamento de categorias temáticas atualmente pequenas, como é o caso do grupo 5, onde podem estar surgindo novas frentes de pesquisa, como por exemplo, a categoria temática de conversão de energia ou aceleradores e outras merecem acompanhamento e são temas em que as ações de política científica podem intervir, no sentido de expandir as pesquisas dependendo das diretrizes políticas nacionais.

5.3 Resultados da verificação dos artigos

Esta parte dos resultados se constitui em um dos objetivos desta tese. Todos os resultados foram obtidos a partir do estudo realizado com os artigos produzidos pelas teses e coletados no Currículo Lattes e se descreverão: a) os resultados da consulta aos autores das teses sobre a correlação dos artigos, para efeito de validação da técnica de coocorrência; b) os resultados da aplicação da técnica comparando com os resultados da opinião dos autores; c) os resultados da aplicação da técnica a todo o universo de artigos; d) o cálculo de intervalo de confiança.

Após todos os procedimentos de coleta já descritos na metodologia deste trabalho, iniciou-se o trabalho com os 2.211 artigos recuperados, estes foram distribuídos com todos os seus dados em matrizes, numerados e codificados de tal forma que permanecessem agrupados com suas respectivas teses. Foi a partir dessa matriz que se identificou a correlação entre as palavras-chave da tese e seus respectivos artigos (Apêndice B).

5.3.1 Resultados da consulta aos autores das teses

Concomitantemente, nesta mesma matriz foram selecionados, de forma aleatória, 128 autores para responderem aos questionários, dos quais 100 (79%) responderam até a época de fechamento da coleta dos dados, analisando um total de 397 artigos (18% do total de 2211). Nesse mesmo universo foi aplicada a técnica de coocorrência de PCs e comparou-se com os resultados obtidos na tabulação das respostas dos autores. Isto significa que após receber a opinião dos autores este universo de 397 itens foi analisado detalhadamente para observar quais eram as características, em nível de indexação, daqueles artigos que foram considerados correlacionados pelo autor. Estas características foram identificadas por meio da técnica de coocorrência de palavras-chave.

Com relação ao questionário (Apêndice A), nesta etapa da análise, com o objetivo de evitar dispersão na resposta dos autores, decidiu-se considerar as respostas 1 e 2 positivas em maior ou menor grau, isto é, denota-se que os artigos **são correlacionados**; portanto, foram agrupadas.

As respostas 3 e 4 foram consideradas negativas, ou os artigos **não são considerados correlacionados**, e também foram agrupadas; assim, trabalhou-se com dois tipos de resposta, **SIM** (questões 1 e 2) e **NÃO** (questões 3 e 4).

Na FIG. 22, se apresentam os resultados das respostas dos autores às quatro alternativas do questionário, indicando a correlação dos artigos em percentual; números da primeira coluna (1, 2, 3,4, em azul).

	Correlação autor	Nº artgs.	%	% acum.
SIM	1	159	40,1	40,1
	2	64	16,1	56,2
NÃO	3	57	14,4	70,5
	4	117	29,5	100,0

Figura 22 – Resumo, em porcentagem, dos resultados das respostas da opinião dos autores com relação à correlação dos artigos

Em uma primeira etapa e para validar a metodologia, verificou-se tanto a coincidência **positiva** como a **negativa** na presunção de que, **se** o autor diz que há correlação entre o artigo e a tese, **então** a indexação deve concordar e apresentar um grau significativo de coocorrência de PCs (FIG. 23, coluna 1, números em vermelho), porém, se o autor diz que **não** há correlação, a indexação também **não** deve apresentar uma incidência significativa de coocorrência de PCs (FIG. 23, coluna 3, números em vermelho).

Foi efetuada então a comparação entre o resultado obtido pelo método automático (indexação com as palavras-chave) e o resultado obtido tabulando a opinião do autor. Nessa comparação, obtiveram-se os seguintes resultados:

Coluna 1 ↓	Coluna 2 ↓	Coluna 3 ↓	Coluna 4 ↓	
Nº de coocorrências de PCs	Correlação Autor (questionário)			
	1	2	3	4
	SIM		NÃO	
0	3	2	18	93
1	4	3	14	9
3	29	12	12	9
6	34	18	10	4
9	1			
10	45	15	2	1
14	1			
15	25	10	1	
21	12	2		1
27	1			
28	4	2		
Total	159	64	57	117

Figura 23 – Resultados comparativos entre o teste automático de coocorrência de palavras-chave e os questionários respondidos pelo autor

Na FIG. 23, se apresenta o resultado comparativo entre as respostas dos autores das teses obtidas via questionários (Apêndice A) e o resultado obtido via método automático. Na coluna 1, observa-se o número de **pares** de palavras-chave

coocorrentes na indexação dos artigos coincidentes com as teses, identificados pelo método automático na amostra dos 397 artigos. Nas colunas 2 e 3, está registrado o número de artigos em que o autor assinala a correlação, denominada SIM e NÃO (respostas extraídas do questionário), e verifica-se que:

- ✓ quando o autor afirma que os artigos têm correlação com a tese, há maior número de coocorrências de PCs em maior número de artigos ou, em outras palavras, a indexação concorda (FIG. 23, coluna **2**, números em vermelho);
- ✓ quando o autor afirma que os artigos **não** têm correlação com a tese, a indexação também concorda; detecta-se a coocorrência de poucas PCs em um número maior de artigos (FIG. 23 coluna **3**, números em vermelho);
- ✓ a porcentagem total de coincidência (itens **a** e **b**) alcança **86,9%**; considerou-se esse resultado adequado para validar a metodologia.

Na FIG. 24, observa-se que, com base na amostra, o método automático considera correlacionados **251** artigos (63,2%), desses, **40** artigos (15,9%) apresentam discordância entre o método (SIM) e a opinião do autor (NÃO) **para um índice de coocorrência de 3 PCs ou mais**.

Existência de correlação pelo método automático	Correlação Autor				Total
	1	2	3	4	
	SIM		NÃO		
Não (0 ou 1 PCs)	7	5	32	102	146(artigos)
Sim (3 ou + PCs)	152	59	25	15	251(artigos)

Figura 24 – Resumo dos resultados comparativos entre o teste automático e a opinião do autor

A discordância entre esses 40 artigos (em vermelho, FIG. 24) é bastante aceitável, uma vez que em um serviço de informação, no processo de indexação dos itens o foco é a recuperação da informação; tomando, por exemplo, um artigo que trate do assunto Irradiação com laser, a frequência da palavra-chave laser será uma constante na indexação, independentemente de o artigo ser correlacionado com a tese do mesmo autor ou não.

Assim, estabeleceu-se que a partir de **3** ou mais coocorrências de PCs entre a tese e o artigo, a correlação é mais alta.

Os resultados obtidos pela correlação entre a opinião do autor e o método automático nessa amostra serviram de base para se estabelecerem os critérios de correlação a serem aplicados no universo total de artigos (2.211). Na literatura sobre *co-word analysis*, não se identificou um valor já estabelecido sobre o número mínimo ou máximo de coocorrências; os autores comentam que o corte nos valores está relacionado à modalidade e ao objeto de estudo.

No caso deste trabalho, o número mínimo de coocorrências identificado pelo método automático foi de 1 (uma) para 7 documentos que o autor disse SIM. Para validação do método 1 (uma) coocorrência, não foi considerada representativa e, procurando ajustar melhor a metodologia, foi adotado o valor de 3 ou mais coocorrências para se afirmar que há uma correlação mais alta.

Já na literatura internacional, conforme citado anteriormente, há um número significativo de trabalhos, principalmente na Europa onde a grande maioria dos trabalhos descritos inicia-se pela extração e preparação dos dados, confecção de matrizes de coocorrência de palavras para depois produzir *clusters*, mapas de conhecimento ou redes sociais, e sempre utilizando as bases de dados já prontas como é o caso da *Web of Science* ou a *Scopus*; não se identificaram trabalhos em que a técnica fosse utilizada para identificar a produtividade das teses.

5.3.2 Cálculo do intervalo de confiança

Objetivando-se assegurar confiabilidade ao trabalho, recorreu-se à assessoria de um profissional da área, para o tratamento estatístico dos dados coletados.

Nessa etapa do trabalho (coocorrência de PCs) especificamente, por se tratar de uma técnica que não foi utilizada, com a mesma finalidade e com o mesmo objeto de estudo, optou-se por aplicar um teste de cálculo de erro amostral nos resultados do método automático e a opinião dos autores.

De acordo com a FIG. 22, quadro de distribuição da correlação dos artigos, pode-se afirmar que há concordância entre a opinião do autor e o método automático quando:

- ✓ ambos afirmam haver correlação (resposta 1 ou 2 do autor) em pelo menos 3 coocorrências de PCs coincidentes; ou
- ✓ ambos afirmam não haver correlação (resposta 3 ou 4 em menos que 3 coocorrências coincidentes).

A amostragem teve como unidade elementar **o artigo** publicado e como unidade amostral **o autor**. Dessa forma, optou-se por trabalhar com uma amostragem por conglomerados.

Segundo BOLFARINE e BUSSAB (2005), o que caracteriza o planejamento amostral por conglomerados é que este contém mais de um elemento populacional. Sortear os artigos poderia aumentar o número de autores a serem consultados; por isso se decidiu sortear os autores e questioná-los sobre sua produção, isto é, sobre os artigos que eles publicaram no período estabelecido para o estudo (5 anos antes e 5 depois).

Dessa forma, define-se a população dos artigos

$$U = \{a_1, a_2, \dots, a_N\}, \text{ onde } N = 2.211$$

que também pode ser representada pelos autores

$$U = \{w_1, w_2, \dots, w_A\}, \text{ onde } A = 401$$

ou suas respectivas produções

$$U = \{w_{1B_1}, w_{2B_2}, \dots, w_{nB_A}\},$$

Onde:

B é o tamanho do conglomerado ou o número de artigos publicados pelo autor, para o cálculo de proporção de concordância foi considerada a amostra de 100 autores cujas produções somaram 397 artigos, onde:

- $a_{ji} = 1$, se o artigo i do autor j apresenta concordância entre a opinião dele e o método automático, e
- $a_{ji} = 0$, caso contrário.

Onde: $n = 100$, $i = 1, 2, \dots, n$ e $j = b_1, b_2, \dots, b_n$.

seja

$c_j = \sum_{i=1}^{b_j} a_{ji}$ o número de artigos do autor j no(s) qual(is) há concordância com o método automático.

Então a proporção de concordância é dada por:

$$p = \frac{\sum_{j=1}^n c_j}{n} = \frac{345}{397} = 0,8690$$

e o respectivo erro padrão é dado pela seguinte fórmula:

$$\sigma(p) = \sqrt{\frac{1}{n.(n-1).b} \sum_{j=1}^n (c_j - p.b_j)^2} = \sqrt{\frac{1}{100.(100-1).3,9700^2} .55,4127} = 0,0375$$

O intervalo de confiança (95%) é dado por:

$$\begin{aligned} & [p - z_{95\%} . \sigma(p); p + z_{95\%} . \sigma(p)] = [0,8690 - 1,96.0,0375; 0,8690 + 1,96.0,0375] \\ & = [0,7954; 0,9426] \end{aligned}$$

Ou seja, o erro amostral de aproximadamente 7,36% resulta num intervalo de 95% de confiança, com limites de aproximadamente 79,5 e 94,3%.

Esse resultado pode ser interpretado afirmando-se que, de todas as possíveis amostras que tivessem sido sorteadas da população de autores, em 95% delas a proporção de concordância estaria compreendida dentro desse intervalo.

Conclui-se que o grau de concordância do método automático com a opinião do autor tem muito pouca probabilidade de ser inferior a 79,5%, validando portanto o método automático, por contagem de coocorrência coincidentes, que passa a ser aplicado à população de autores.

5.3.3 Aplicação do método de coocorrência no universo total de artigos

Após essa validação inicial, retomou-se o número total de artigos, 2.211, e aplicando-se os mesmos critérios obtidos com a amostra (método automático), é possível observar que a correlação mais forte do artigo com a tese de origem se dá quando há a ocorrência de **três pares** ou mais de palavras-chave. Assim, obteve-se um total de 1.237 artigos considerados **correlacionados**, isto é, produtos da tese propriamente ditos. Na FIG. 25 apresentam-se esses resultados.

Correlação	Total de artigos	%
<i>Nº de coocorrências</i>	<i>Frequência/Nº artigos</i>	<i>%</i>
Não	974	44,1
0	670	30,3
1	304	13,8
Sim	1237	55,9
3 a 10	965	43,6
11 a 20	165	7,5
20 ou mais	107	4,8
Total	2211	100

Figura 25 – Resultados obtidos, em percentual, após a aplicação do teste automático no universo total de artigos (2211) 2, produzidos pelas teses no período analisado 1982-2009

Objetivando apresentar um procedimento mais detalhado sobre o uso da técnica de correlação ao universo total de artigos, faz-se necessário comentar cada etapa, assim volta-se ao início do procedimento do trabalho realizado com os artigos coletados.

Vale lembrar também que para a identificação da produção de artigos foram consideradas 401 teses, ou seja, do universo total de teses (536) foram excluídas aquelas que não **registraram nenhum artigo**, como, por exemplo, as teses correspondentes a 1977 até 1981 (12), e assim por diante nos anos subsequentes. Observe-se que na TAB. 4 a data de início de publicação das teses, para esta etapa da análise, é 1982. A opção de trabalhar somente com as teses cujos autores declararam ter publicado artigos no CL, se deve, principalmente, ao fato de ser imprescindível dispor da evidência física para realizar o processo de indexação e posteriormente aplicar a técnica de coocorrência.

Na FIG 26, apresentam-se os resultados do levantamento inicial no Currículo Lattes, do número de artigos que cada tese produziu respectivamente distribuídos pelo ano de publicação da tese e pelo ano de publicação dos artigos, bem como se apresenta uma linha de tendência para evidenciar a evolução destes.

Chama-se atenção para a leitura desses dados. Trata-se da apresentação dos resultados de um levantamento numérico inicial, ou “dados brutos”, sem nenhum tratamento ainda. São resultados similares aos que são frequentemente, apresentados por estudos sobre o número de artigos produzidos pelas teses, como, por exemplo, a FIG. 5 apresentada por PENA (2010), para demonstrar o crescimento da produção de artigos, p. 45 deste trabalho. Esses dados apresentados dessa forma fornecem um determinado nível de informação, porém, não respondem às questões que a análise deste estudo pretende identificar, como, por exemplo, qual é o nível de correlação de cada artigo com a tese? Sem a elaboração da tese, qual seria o número de artigos publicados? Identificando-se esta correlação é possível afirmar com maior segurança qual a contribuição de cada tese, em nível de publicação de artigos.

Realizando-se uma comparação entre as duas FIG. 5 e 26 (p. 45 e p. 128), observa-se que as linhas que informam a quantidade de artigos produzidos são bastante semelhantes, o que é justificável, uma vez que o Programa de Pós-graduação do IPEN se integra ao contexto nacional de pós-graduação, e coincidentemente apresenta uma mesma tendência.

Figura 26 – Distribuição temporal dos artigos, pelo ano de publicação da tese e pelo ano de publicação do artigo, sem a aplicação da técnica de correlação

Na TAB. 5 são apresentados os dados coletados no Currículo Lattes distribuídos por ano de defesa da tese, ainda sem a aplicação da técnica de coocorrência. Um fato importante a se considerar na leitura desses dados é que o ano de publicação da tese, na maioria das vezes, difere do ano de publicação do artigo, isto se deve à metodologia de coleta adotada, qual seja, considerar os artigos publicados no período de 5 anos antes e 5 anos depois da data de defesa da tese

Exemplo: o autor de uma tese apresentada em 2002 pode não ter publicado nenhum artigo em 2002, e sim em 2001 ou 2003, e assim sucessivamente. Para fins de análise destes dados, procurou-se trabalhar sempre com a data de defesa da tese, uma vez que o principal objetivo nesta etapa do trabalho foi identificar a produtividade das teses.

Tabela 5 – Distribuição das teses e dos artigos publicados e declarados pelos autores das teses no Currículo Lattes, distribuídos por *ano de defesa da tese*, sem a aplicação da técnica de coocorrência.

Ano de defesa da tese	Nº de Teses	Nº de Artigos
1982	3	8
1983	3	10
1984	5	16
1985	4	14
1986	2	2
1987	1	2
1988	5	26
1989	2	11
1990	3	12
1991	2	11
1992	5	30
1993	10	63
1994	12	62
1995	10	61
1996	11	91
1997	17	57
1998	14	79
1999	24	97
2000	23	146
2001	23	220
2002	15	72
2003	18	118
2004	30	266
2005	41	218
2006	40	188
2007	22	141
2008	32	115
2009	24	75
Total geral	401	2211

Tabela 6 – Distribuição temporal dos artigos declarados pelos autores das teses, coletados no Currículo Lattes, sem a aplicação da técnica de coocorrência

Ano de publicação do artigo	Nº de Artigos
1977	2
1979	1
1980	5
1982	3
1983	4
1984	6
1985	7
1986	17
1987	10
1988	14
1989	19
1990	16
1991	16
1992	21
1993	28
1994	43
1995	64
1996	66
1997	60
1998	80
1999	94
2000	88
2001	143
2002	204
2003	168
2004	210
2005	209
2006	224
2007	136
2008	149
2009	95
2010	9
Total geral	2211

Na TAB. 6 se apresenta a distribuição temporal dos artigos coletados do Currículo Lattes declarados pelos autores das teses e distribuídos pelo **ano de publicação do artigo** e ainda sem a aplicação da técnica de correlação.

Observe-se que embora o total seja igual ao da TAB. 5 a distribuição temporal difere da distribuição das teses, registram-se assim artigos desde 1977 até 2010; isto se deve, conforme já foi comentado, ao período estabelecido para coleta dos artigos.

Por exemplo, uma tese defendida em 2008 pode gerar artigos publicados correlacionados em 2007, 2009 ou 2010, porém não no mesmo ano de defesa da tese.

Por outro lado, em um primeiro olhar na TAB. 6, vê-se que o ano de 2006 registra o maior numero de artigos (224), isto não quer dizer que as teses de 2006 geraram todos estes artigos publicados, o ano de publicação do artigo está correto e certamente houve uma concentração de publicações, porém as teses que geraram esses artigos podem ter sido defendidas antes ou depois do ano de 2006.

Na TAB. 7, apresentam-se de forma resumida os resultados obtidos após a aplicação da técnica a todo o universo de artigos, isto é, os 2.211. Para facilitar a visualização, optou-se pela distribuição quinquenal das 401 teses. Apresenta-se também o número de artigos produzidos e declarados pelos autores das teses no CL, bem como os artigos correlacionados com as teses, identificados pela aplicação do método automático, os artigos correlacionados publicados em revistas internacionais e indexadas na Web of Science (WoS).

Com o objetivo de se homogeneizarem os períodos de análise e obter uma visão da média de artigos produzidos, com outro cenário de tempo, e considerando que as teses produzidas após 2005 não teriam tempo suficiente para produção total de artigos, que interferisse nos resultados, foi selecionado um período intermediário de 20 anos, de 1985 a 2004 (TAB. 7), indicado em cor amarela.

Tabela 7 – Distribuição das teses, do nº total de artigos declarados pelos autores das teses e o nº de artigos correlacionados publicados em periódicos internacionais e indexados na WoS. A cor amarela evidencia o período intermediário selecionado, mostrando os resultados sem interferência do período estabelecido na coleta dos dados (5 anos antes e 5 anos depois)

Quinquênio tese	Nº de teses	Nº de artigos total	artigos Corr.	% de Corr.	Corr. em PI	% de Corr. em PI	indexados na WoS
1982-1984*	11	34	16	47,1	13	35,3	12
1985-1989	14	55	22	40	13	18,2	10
1990-1994	32	178	99	55,6	58	29,8	53
1995-1999	76	385	211	54,8	154	33	127
2000-2004	109	822	482	58,6	387	33,6	276
2005-2009	159	737	407	55,9	341	32,6	242
Total geral	401	2.211	1.237		966		720

* adotado como triênio para facilitar a distribuição dos demais períodos

– Legenda: **Corr.** - Correlacionados; **PI**- Periódicos Internacionais; **WoS** - Web of Science

Na TAB. 8 é possível observar que no período intermediário selecionado 1985-2004 buscando evitar viés; há diferença entre a “média total” a qual é mais alta (sem filtro de correlação), e a média obtida após a aplicação da técnica de correlação, esta última reflete de forma mais real a produtividade, tornando os indicadores mais representativos nesse aspecto.

Tabela 8 – Nº de teses, média do nº total de artigos, e média do número de artigos correlacionados no período intermediário analisado 1985-2004

Quinquênio tese	Nº de teses	Média Total	Média Correlacionados
1985-1989	14	3,9	2,0
1990-1994	32	5,6	3,7
1995-1999	76	5,1	3,4
2000-2004	109	7,5	4,8

Tabela 9 – Média total de artigos produzidos e declarados pelos autores das teses e média de artigos correlacionados com a tese após a aplicação da técnica de coocorrência, distribuídos pelo ano de defesa da tese, período 1982-2009

	Coluna A	Coluna B	Coluna C
Ano da tese		Média de artigos por tese	
	Total	Correlacionados	Corr. em PI
1982	2,7	2,0	2,0
1983	3,3	1,7	1,3
1984	3,2	2,3	2,3
1985	3,5	1,7	1,3
1986	1,0	1,0	1,0
1987	2,0	0	0
1988	5,2	1,8	1,3
1989	5,5	4,0	1,5
1990	4,0	3,5	5,0
1992	6,0	3,8	1,6
1993	6,3	4,4	4,3
1994	5,2	3,0	1,7
1995	6,1	3,7	5,3
1996	8,3	6,9	3,7
1997	3,4	2,6	2,3
1998	5,6	3,6	3,3
1999	4,0	2,6	2,4
2000	6,3	4,6	4,1
2001	9,6	6,3	5,5
2002	4,8	3,8	3,4
2003	6,6	3,6	3,2
2004	8,9	5,1	4,2
2005	5,3	2,9	2,5
2006	4,7	3,2	2,8
2007	6,4	4,2	3,7
2008	3,6	2,4	2,5
2009	3,1	2,6	2,8

Na TAB. 9 se apresenta a média total de artigos publicados e declarados pelos autores das teses, discriminados por ano no período 1982- 2009, para demonstrar a diferença entre resultados obtidos calculando uma média simples (Coluna A) e a média obtida após a aplicação da técnica de correlação (Coluna B). A técnica de correlação permitiu auferir uma média mais real de artigos publicados a partir da elaboração da tese.

Observa-se também que a media de artigos varia de acordo com o ano, o qual não permite generalizações, bem como a média dos artigos publicados em periódicos internacionais na coluna C.

Na FIG. 27 visualiza-se a evolução dos mesmos dados ao longo do período estudado. Observe-se que no ano de 1991 há uma interrupção na linha do gráfico, o que se deve ao fato de que, embora tenha havido artigos publicados declarados pelos autores das teses, não se identificaram artigos correlacionados nesse período, após a aplicação da técnica de coocorrência.

Figura 27 – Distribuição do nº total de artigos, correlacionados, publicados em periódicos internacionais, e indexados na WoS no período estudado, 1982-2009

Figura 28 – Distribuição quinquenal comparativa dos resultados, evidenciando o nº de teses, a diferença entre o nº total de artigos e o nº de artigos correlacionados, e a linha de tendência destes itens

5.3.4 Resultado da verificação do tempo para publicação dos artigos

Um aspecto analisado importante, também verificado nesta etapa (registrado nos objetivos específicos do trabalho), foi a verificação do tempo em anos para publicação dos artigos a partir da data de defesa da tese.

Na FIG. 29 constata-se, a partir de 2004, uma diminuição acentuada e sequente, na média de tempo, para publicação dos **artigos correlacionados**, alcançando em 2009 a média de 1,63 ano, **antes** da defesa da tese. Trata-se de fato positivo para o processo de comunicação científica, porém este dado deve ser visto com cautela; é provável que as teses defendidas mais recentemente, por exemplo, nos últimos quatro anos (assinalados na figura com linha pontilhada), não tenham tido tempo suficiente para a publicação de artigos. Somente um estudo realizado nos próximos cinco anos poderá corroborar esta tendência.

Figura 28 – Média de tempo, em anos, para publicação dos artigos correlacionados (1237), em relação à data de defesa da tese

6. DISCUSSÃO

6.1 A análise das séries temporais

O emprego de indicadores bibliométricos apresenta uma série de vantagens diante de outros métodos utilizados na avaliação científica. Por se tratar de dados numéricos, verificáveis, reproduzíveis, eles apresentam resultados objetivos e podem ser aplicados a um grande volume de dados.

Os resultados devem ser úteis para os gestores da Instituição; devem sustentar suas decisões sobre a política científica, confirmar decisões adotadas, fundamentar correções de rumo, ou ainda criar *agendas* de pesquisa futuras. Por meio da análise histórica foi possível observar como o desempenho do Programa está condicionado ao estilo de gerenciamento dos gestores da Instituição.

As temáticas abordadas nas dissertações e teses mostraram o histórico da trajetória de um Programa bem-sucedido, operando em uma área claramente institucionalizada social e cognitivamente, bem estruturado e abrangente, permitindo a adequação necessária à demanda da sociedade ao longo do tempo; não é um Programa rígido, está sempre atento às mudanças, porém, sem perder a percepção de sua missão. A crescente interdisciplinaridade da ciência é uma característica marcante no estágio de globalização atualmente vivenciado; esse fato demanda a necessidade de se acompanhar a dinâmica de produção de conhecimentos.

No caso dos resultados numéricos das dissertações e teses (FIG. 13 p. 107 deste trabalho), há ainda uma peculiaridade devido ao fato de se estar trabalhando com esse tipo de objeto, que demanda um determinado período de tempo para sua elaboração. Os totais anuais mostrados são sempre resultado de ações ou decisões adotadas no mínimo quatro a seis anos antes, prazo de confecção da dissertação ou da tese. Tomem-se, por exemplo, os números registrados logo no início do período, ano de 1978, quando apareceram 51 dissertações e teses; certamente foram trabalhos iniciados fora do IPEN, vindo a ser apresentados no Programa da PG após seu início.

Observa-se, também, a influência das questões políticas interferindo na produção da ciência, nos anos 1986 e 1987, quando o Programa apresentou um número muito reduzido de trabalhos, resultado de uma decisão política adotada em 1982 pela gestão do então Superintendente do IPEN, com o intuito de desativar o Programa de PG. Não havendo matrículas, quatro ou cinco anos depois não havia produção. Celebra-se hoje a obstinação dos pesquisadores da época em manter o Programa mesmo contrariando decisões superiores (VIEIRA; AQUINO e ROGERO, 2010)²⁹. A não ser por esse fato significativamente marcante, nos anos de 1977 a 1992, observa-se um crescimento intermitente com uma média discreta de produção, ainda que nesse período tenha havido um empenho governamental com a criação de um programa voltado especialmente para a formação de recursos humanos, o PRONUCLEAR.

Em 1975, por meio da Portaria Interministerial n. 93, de 8 de setembro, o governo criou um grupo de trabalho com a atribuição de propor um programa de formação de recursos humanos que pudesse atuar dentro do recém-assinado Acordo Nuclear Brasil-Alemanha Ocidental sobre cooperação no campo dos *Usos Pacíficos de Energia Nuclear*. Assim, foi criado o Programa de Recursos Humanos para o setor nuclear (PRONUCLEAR) em julho de 1977 (SCHMIEDECKE e PORTO, 2008). O nome do programa é autoexplicativo, e os alunos recebiam bolsas de estudo, além de o programa subvencionar as instituições participantes. Além de cursos de especialização para os profissionais, foram selecionados cursos de pós-graduação no país e no exterior, entre eles, o Programa de PG do IPEN para a formação de Mestres.

No período de 1977 a 1986, 131 bolsistas completaram o curso de Mestrado no IPEN; desse grupo de bolsistas aproximadamente 35 foram enviados para o exterior a fim de realizarem o doutoramento ainda dentro do programa PRONUCLEAR. Em 1983, o programa foi desacelerado, e finalmente desativado em 1986 (PRONUCLEAR, 1986).

²⁹ Comunicação pessoal fornecida por VIEIRA M. M.; AQUINO A. e ROGERO, J. R., em São Paulo, 2010.

A partir de 1992, novamente por decisão política (GORDON, 2003) desta vez favorável, deu-se a abertura para novas áreas de pesquisa, e o Programa iniciou a sua linha ascendente, com alguns períodos de queda, mas reagindo no período seguinte, motivado por decisões gerenciais de ajuste e readequação, como por exemplo, a decisão de se aplicar um exame de inglês mais rigoroso como meio de seleção para admissão no Programa.

Em 1997, as áreas de concentração do Programa de Pós-graduação foram readequadas e, em vez de duas, subdividiram-se em três; porém, durante algum tempo as duas situações ainda conviveram, como era de esperar (INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES, 2005). Houve períodos de alta numérica expressiva, como os anos de 2005 e 2006 (FIG.13), que poderiam ser creditados como refluxo no ajuste do Programa devido à decisão, por exemplo, em 2000, de reduzir o prazo máximo para a finalização dos trabalhos, ou à decisão de não mais prorrogar os prazos para aqueles que estivessem em processo de elaboração, ou ainda ao incremento da quota de bolsas outorgadas pelo Conselho Nacional de Pesquisas (CNPq).

Em julho de 1993, o Governo Federal criou a Lei n. 8691 (BRASIL, 1993) sobre o Plano de Carreiras dos funcionários públicos fixando critérios para a concessão de acréscimos salariais por titulação; este fato abriu perspectivas na carreira de pesquisador, que se encontrava desmotivado e com os vencimentos achatados por falta de uma política clara e condizente. Uma boa parcela dos funcionários da Instituição decidiu completar ou mesmo iniciar seu ciclo acadêmico. Pode-se observar na FIG. 13, a partir do período de 1999, o reflexo dessa decisão não só no número de teses como também no de dissertações. Este é um exemplo prático de como as decisões políticas podem influenciar na produção de ciência.

Em 1999, os gestores do Programa, atentos à demanda da sociedade, sentiram necessidade de desenvolver trabalhos com a finalidade de instrumentalizar alunos na área de gestão; foi tomada a decisão de se criar uma linha de pesquisa em Gestão, desta vez em convênio com os pesquisadores da Faculdade de Economia e Administração (FEA/USP). Assim, 47 dissertações e teses foram apresentadas até 2006, quando o convênio foi suspenso.

O bom desempenho dos números apresentados pode também ser creditado ao fato exitoso de se alcançar nível 6 na avaliação da CAPES em 2001 (INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES, 2005).

Desde sua criação o Programa tem mostrado qualidade e resultados positivos, sempre avaliado com nível A e continuando nesse patamar até os dias de hoje, refletindo o mérito dos gestores do Programa, a maioria funcionários de carreira da Instituição, pesquisadores, com grande empenho na condução do Programa, cientes da importância da sua manutenção na formação de recursos humanos especializados (ROGERO, 2009)³⁰. Vale lembrar que no país não há muitos cursos de PG similares; para muitos alunos, esse Programa de Pós-graduação é o primeiro contato com a área nuclear.

A distribuição temática (FIG. 14) registra a dinâmica dos domínios científicos da Instituição; são como “espelhos” que refletem as atividades de pesquisa que se desenvolveram em determinadas épocas. Nas FIG. 17, 18, 19, se apresenta a evolução temática ao longo do período estudado. No primeiro período estudado, observam-se 17 áreas, com grande ênfase para os trabalhos em Química (devido ao projeto de desenvolvimento do ciclo de combustível, ver comentário na p. 141 deste trabalho). Já no período seguinte (FIG. 18), em que aconteceu a interferência política da Direção do Instituto, não só o número de trabalhos se reduz como também o número de categorias, este foi considerado o período mais desfavorável para o funcionamento do Programa de PG (VIERA; AQUINO, ROGERO, 2010)³¹. Pode-se observar também, pelo número de trabalhos, o fortalecimento do Programa a partir de 1994. Durante o período estudado foram se apresentando outras categorias temáticas, como por exemplo, a de conversão de energia e aceleradores cujos trabalhos começam a aparecer a partir do quinquênio 1995-1999, ainda que em pequena quantidade. Atualmente, os trabalhos se distribuem em 21 categorias, confirmando a multidisciplinaridade da área nuclear. Somente um estudo posterior demonstrará a sazonalidade das categorias emergentes.

³⁰ Comunicação pessoal fornecida por ROGERO, J. R., em São Paulo, 2009.

³¹ Comunicação pessoal fornecida por VIEIRA M. M.; AQUINO A. e ROGERO, J. R., em São Paulo, 2010.

O estudo de séries temporais é elucidativo, porém mostra uma situação em determinado momento; a produção dos dados tem que ser contínua, e os estudos, sequentes; somente assim, o processo se consolida e adquire confiabilidade como instrumento de gestão.

Os indicadores temáticos fornecem uma variedade de informações; dependendo do agrupamento adotado, permitem interpretações diferentes; por isso se optou por apresentar os mesmos dados com agrupamentos diferentes.

6.2 A correlação dos temas com a história do Programa de Pós-graduação do IPEN

Desde a década de 1960, ou seja, antes do início do Programa, em 1977, até o fim da década de 1980, quando foi anunciado o domínio do ciclo do combustível, a Instituição estava empenhada em desenvolver pesquisas e atividades relacionadas ao ciclo do combustível nuclear³², uma tecnologia de ponta que envolve processos químicos complexos.

O apoio político e financeiro era favorável para que a Instituição realmente dominasse o ciclo de enriquecimento do combustível nuclear. Havia uma equipe já atuante composta de pesquisadores seniores dedicados às atividades de purificação de elementos químicos e obtenção de compostos de urânio nuclearmente puros LIMA (1961); ABRÃO (1962); GORDON (2003). Ainda nesse período, desenvolveram-se as pesquisas sobre o reprocessamento e a separação isotópica.

Em 1977, instalou-se a CELESTE I, planta piloto de reprocessamento de material radioativo, e em 1979, iniciou-se o **Projeto Conversão** (PROCON), destinado à produção de hexafluoreto de urânio (GORDON, 2003). Essas atividades estão “registradas” na distribuição temática das dissertações e teses defendidas

³² O Ciclo do Combustível Nuclear pode ser entendido como um conjunto de processos realizados em plantas de tratamento químico do urânio onde se desenvolvem diversas operações unitárias que, direta ou indiretamente, promovem diferentes graus de purificação do urânio em cada uma das etapas, segundo condições físico-químicas controladas, até alcançar o grau adequado para seu uso como combustível nuclear. O ciclo do combustível também abrange o reaproveitamento do urânio não consumido no reator nuclear. Comunicação pessoal, fornecida por AQUINO, A., em São Paulo, 2010.

nesse período (FIG. 17 p. 112 neste trabalho), principalmente entre 1977 e 1984, quando se destacaram os trabalhos na área temática em química analítica. Nos cinco anos subsequentes, as atividades nessa área se dedicaram exclusivamente às operações do domínio do ciclo do combustível e culminaram com o anúncio desse fato pelo então presidente do Brasil, José Sarney, em 4 de setembro de 1987 (MARQUES, 1992).

A partir da década de 1990, por decisão política de âmbito nacional, as pesquisas nessa área foram praticamente encerradas. A partir dessa data, os integrantes da área canalizaram suas atividades para o desenvolvimento de tecnologia química e meio ambiente visando sua aplicação na área nuclear, saúde e indústria, com sucesso (AQUINO, 2010)³³.

Grande parte dos então alunos do Programa se tornou pesquisadores, dividindo suas funções entre a pesquisa e a atividade acadêmica.

A análise de trajetória mostra que a área superou a mudança de atividades, apresentando, a partir do terceiro quinquênio, produção acadêmica com tendência à elevação (FIG. 21, G2, p.116).

Outra área temática que se destaca pelo desempenho é a Ciência dos Materiais, criada como Programa de Ciência e Tecnologia de Materiais, no IPEN em 1971. A exemplo da área de Química, na época do início do Programa de PG, nesta área também já se desenvolviam atividades direcionadas à produção dos elementos combustíveis experimentais para reatores, sob a coordenação de pesquisadores seniores (SANTOS, citado por GORDON, 2003). É provável que o fato de já existir uma competência instalada nessa área tenha influenciado o momento da criação do Programa de PG (1976), uma vez que foi iniciado com duas áreas de concentração: Reatores Nucleares de Potência e Tecnologia do Combustível, e Tecnologia Nuclear Básica.

Nesse grupo, também havia uma concentração de ex-bolsistas do programa PRONUCLEAR que realizaram o doutoramento fora do país; ao retornarem,

³³ Comunicação pessoal, fornecida por AQUINO, A., em São Paulo, 2010.

encontraram uma situação adversa devido à mudança nos rumos da política nuclear, e tiveram que redirecionar suas atividades, incluindo a orientação de alunos de PG e a apresentação de projetos para os órgãos de fomento; com resultados positivos, criaram-se oportunidades para novas linhas de pesquisa, obtenção de bolsas, laboratórios e, conseqüentemente, maior produção científica (BRESSIANI, 2008)³⁴. Os resultados do FIG16 (p.111) espelham essa dinâmica; a área de Ciência dos Materiais responde por 24% do volume de dissertações e teses em comparação com o grupo das nucleares; desde sua criação, teve e continua apresentando uma atuação muito intensa; os pesquisadores dividem seu tempo entre as atividades de pesquisa e de formação de pessoal, e desde o segundo quinquênio, observa-se elevada tendência a crescimento (FIG.20, G1 p.116).

A área de tecnologia de reatores esteve sempre presente, como era de esperar em uma Instituição com um Programa de PG criado com essa finalidade, embora com períodos de produção intermitente, novamente fruto da política de atuação vigente na Instituição; exemplo disso pode ser observado na FIG.20, G2 p.116, em que se percebe claramente uma queda acentuada no segundo quinquênio.

Havia na década de 1980 uma *massa crítica* de conhecimento, formada por profissionais bem preparados, em sua maioria engenheiros ou físicos, ex-bolsistas participantes do Programa PRONUCLEAR, com o doutorado realizado nas principais universidades nos Estados Unidos ou na Europa; ao voltarem ao país, esses profissionais não encontraram uma situação favorável de aproveitamento dos seus conhecimentos devido à política vigente para o setor nuclear na época.

Assim, ou deixaram a Instituição para atuarem em empresas do setor produtivo ou se integraram às atividades de desenvolvimento de projetos para o Programa Nuclear da Marinha (1980 até 1992)³⁵, sem opção para atuarem na área acadêmica.

³⁴ Comunicação pessoal fornecida por BRESSIANI, J. C., em São Paulo, 2008.

³⁵ O Programa da Marinha foi criado em 1979 com o objetivo de desenvolver a propulsão nuclear em submarinos. Para isto construíram-se laboratórios, montaram-se equipes de engenharia e constituíram-se convênios com instituições de pesquisas para desenvolver todas as etapas do ciclo do combustível nuclear (RABELO, 2007).

A partir da década de 1990, com o fim do projeto da Marinha, houve oportunidade de dividir suas atividades com o Programa de Pós-Graduação; formação e conhecimento para isso já haviam sido adquiridos. Atualmente, constitui-se em uma área com forte tendência a crescimento (FIG. 21, G2 p.116).

Outra área que com desempenho significativo desde o início das atividades do Programa é a dedicada a desenvolver estudos sobre os efeitos das radiações, tanto internos como externos, nos organismos vivos, em nível celular e bioquímico; assim, os estudos enfocam os procedimentos de irradiação, doses absorvidas, modificações genéticas, toxicidade, consequências imunológicas e outros, tanto no homem quanto nas plantas e animais (IPEN, 2010)³⁶.

De certa forma, é a área que mais evidencia a multidisciplinaridade da energia nuclear; os autores dos trabalhos são biólogos, médicos, farmacêuticos, veterinários, bioquímicos; muitas das pesquisas desenvolvidas servem de base para importantes aplicações terapêuticas, bem como, de forma geral, têm papel social representativo na área de saúde, agricultura e criação de animais.

No início do Programa, não havia uma área de concentração específica para esse tema; porém, a partir de 1997, os trabalhos que envolviam as aplicações de técnicas nucleares foram agrupados em uma só área de concentração dentro do Programa de Pós-graduação. Trata-se de uma área cujos profissionais, atualmente pesquisadores seniores, dedicam-se tanto à área de pesquisa como à formação de pessoal, mostrando tendência a crescimento (FIG. 21, G3 p.116 deste trabalho).

Os trabalhos na categoria temática de Meio Ambiente apresentam também uma trajetória interessante. Atualmente, o tema está muito difundido e incorporado às práticas cotidianas da sociedade; porém, a linha dessa categoria temática mostra que desde o início do Programa de PG em 1977 esse tema já era abordado em maior ou menor grau nas dissertações e teses denotando a preocupação da instituição nesse aspecto. Os trabalhos abordam pesquisas sobre o uso das técnicas nucleares, tanto para proteger ou conservar o meio ambiente como para monitorar os danos que a energia nuclear pode lhe causar; o fato é que se trata de uma categoria

³⁶ IPEN página eletrônica, 2010: < www.ipen.br >.

cujos trabalhos, a continuarem nesse ritmo, indicam forte tendência a crescimento (FIG. 21, G3 p.116 deste trabalho).

Um olhar mais cuidadoso perceberá as categorias temáticas que, embora no passado não apareçam, nos dois últimos quinquênios iniciaram sua atuação, discreta e não muito rápida; denotam, porém, um potencial de crescimento e merecem um acompanhamento nos próximos períodos.

Exemplo deste fato é a categoria “Energia” e, nesse caso, trata-se da energia não convencional. Seguindo uma tendência mundial, e dentro de uma Diretriz Nacional, em 2000 se iniciou no IPEN uma linha de pesquisa na área de fontes energéticas eficientes e de baixo impacto ambiental, o estudo e desenvolvimento de sistemas associados à tecnologia de Células a Combustível³⁷. Os estudos preliminares já haviam sido iniciados em 1998 na área de materiais valendo-se da experiência desse grupo. Na FIG. 19, é possível observar a presença de dissertações e teses nessa categoria temática a partir de 2005.

6.3 A produtividade das teses com relação aos artigos publicados

Na história da comunicação científica, os periódicos se tornaram o canal de comunicação de excelência, funcionando como uma instância de consagração.

ZIMAN, em 1979, já comentava que sem a comunicação uma área científica não poderá existir, pois o conhecimento resultante das pesquisas não será validado e nem considerado científico.

Os artigos publicados nos diversos periódicos científicos materializam esse fluxo de comunicação e permitem o controle institucionalizado do conhecimento, visto que eles só acontecem após um processo de críticas, aceitação e validação da comunicação científica.

Esse meio de comunicação é tão importante que tem sido utilizado como instrumento de avaliação decisório pelos órgãos de fomento (CAPES, CNPq,

³⁷ IPEN página eletrônica, www.ipen.br Programa Grupo de Células a Combustível, 2000.

FAPESP); eles criaram mecanismos de monitoração constante pelos serviços de informação, tanto nacionais como internacionais, como prova disso a CAPES, o MCT e a FAPESP utilizam os dados extraídos da base WoS para confecção dos indicadores de Ciência e Tecnologia extraídos da base (Índice de citação, Fator h , Fator de impacto dos títulos). O mundo da comunicação científica, no Brasil e no mundo, gravita em torno desses índices.

A elaboração da tese proporciona ao aluno um aprofundamento na construção do conhecimento, tanto no conteúdo como na logística que envolve um trabalho científico; é de esperar que haja prosseguimento nesse processo, que pode acontecer a partir da publicação de um trabalho na forma de artigo.

Ao se efetuar um levantamento do número de artigos produzidos por cada tese defendida dentro do Programa de PG do IPEN, objetivou-se conhecer: **a)** a contribuição de cada tese; **b)** o tempo transcorrido no fluxo de comunicação do trabalho.

Os resultados obtidos na aplicação da metodologia de coocorrência de palavras-chave serviram para efetuar um “filtro” mais adequado para identificar os artigos realmente derivados da tese; dito de outra forma, aqueles artigos que, sem a elaboração da tese, não poderiam ter sido produzidos.

A seleção das palavras-chave como elemento identificador de relação entre os itens mostrou um grau de confiabilidade aceitável. Há que se considerar que a indexação atribuída pelos indexadores carrega um grau de subjetividade inerente ao processo humano, motivo pelo qual alguns autores (CALLON et al., 1991; WHITTAKER, 1989) estudiosos do assunto registraram algumas limitações ao utilizarem a técnica de coocorrência de PCs; entre elas, poderia ocorrer o que eles denominaram o *efeito/indexador*³⁸, a desatualização do vocabulário controlado e a inconsistência dos dados.

³⁸ O *efeito/indexador* se refere ao fato de os profissionais da informação realizarem a indexação manualmente, por meio de um vocabulário controlado e, neste momento de atribuição de conceitos, baseados no seu conhecimento específico, usarem sua subjetividade para selecionar os termos adequados para representar os documentos.

O *efeito/indexador* foi sempre uma preocupação no desempenho da tarefa de indexar a produção científica nessa unidade de informação do IPEN, de onde se obtiveram os dados, muito antes de se cogitar um estudo como este.

Os profissionais encarregados de desempenhar essa tarefa primeiro eram submetidos a um período de treinamento intensivo, tanto na técnica de indexação, no uso do tesauro, como no domínio da área indexada, no caso, a área nuclear; essa prática por certo não elimina totalmente, porém, minimiza o *efeito/indexador*.

Ainda, a consulta aos autores das teses foi uma das ações mais importantes para validar a metodologia utilizada e minimizar o *efeito/indexador*.

Quanto à atualização do Tesauro e à consistência dos dados, como já foi comentado, não procede a preocupação no caso da energia nuclear; a atualização do Tesauro do INIS é anual e a consistência dos dados foi exaustivamente verificada.

Assim, após a aplicação da metodologia, observou-se que 1.237 artigos (55,9%) são correlacionados com as teses (FIG. 25 p. 126 deste trabalho). Desses artigos, se forem considerados somente os correlacionados, 78,1% são publicados em periódicos internacionais, fato altamente positivo para a divulgação e visibilidade dos trabalhos realizados na Instituição.

Se for considerado que a comunidade na área nuclear no país não é tão numerosa por não existirem tantos centros desenvolvendo pesquisa na área nuclear, o que pode ser comprovado pelos poucos Programas de pós-graduação em operação no país, o número de artigos publicados em periódicos internacionais especializados é expressivo; os periódicos são os canais mais apropriados para a divulgação da ciência praticada na Instituição, além de garantirem sua inserção na esfera científica internacional.

Quando se trata de apreciar o mérito da Produção Científica de uma Instituição, de uma área, ou mesmo de um país, um dos critérios mais relevantes é o grau de internacionalização medido pelo desempenho desta na *Web of Science* (índice *h* institucional). A *WoS*, guardadas as peculiaridades, constitui um valioso instrumento de medida e monitoramento da visibilidade institucional.

Mantidas as atuais condições e desempenho do programa de PG na Instituição, as teses continuam a produzir artigos em revistas com visibilidade; seria recomendável, porém, dentro da política de gestão científica, o estabelecimento de mecanismos de incentivo para provocar um esforço da comunidade científica no sentido de avançar não mais na quantidade, mas na busca por elevar o patamar de visibilidade dos trabalhos publicados em revistas internacionais, isto, além de repercutir em maior visibilidade, proporcionará maior reconhecimento por parte da sociedade, tanto da Instituição como dos pesquisadores atuantes.

Estudos longitudinais complementares devem ser efetuados para se obterem parâmetros comparativos que permitam o acompanhamento da evolução.

Um dado significativo diz respeito à média de tempo transcorrido, em anos, para a produção dos artigos correlacionados a partir da apresentação da tese. Observou-se que este tempo está sendo abreviado, alcançando em 2004 uma média de 0,31 ano, para a publicação de um artigo, na FIG. 29 (p.136 deste trabalho) também é possível observar que a partir de 2005 os artigos estão sendo publicados antes da finalização da tese, porém esta constatação não é definitiva e deve ser vista com cautela, uma vez que pode não ter havido tempo hábil para publicação de um número maior de artigos decorrentes das teses defendidas em 2008, 2009, por exemplo, essa tendência deve ser corroborada no próximo quinquênio (FIG.29 p. 136).

Essa constatação de abreviação do tempo para publicação de artigos é significativa principalmente se for analisada pelo aspecto que o Prof. Guimarães, presidente da CAPES, define muito apropriadamente como: *uma atenuação da distância temporal entre a produção de conhecimento e sua apropriação pública* (GUIMARÃES, 2010).

Constitui-se uma obviedade afirmar que esse comportamento está, em parte, fortemente relacionado à evolução e disponibilidade da tecnologia da informação.

A título de exemplo, cita-se a grande mudança ocorrida, e atualmente em prática, em toda a logística operacional necessária para a publicação de um artigo. Embora a prática de avaliação pelos pares (*peer review*) continue inalterada, o fluxo

e o *modus operandi* (desde a produção, elaboração, editoração, envio para o *peer-review* até a concretização da publicação de um artigo) foram totalmente reestruturados graças às facilidades operacionais e à comunicação eletrônica.

Outro fato motivador responsável pelo crescimento do índice de publicação de artigos, bem como pela diminuição do tempo para publicação é a política de incentivo e as diretrizes estabelecidas pelo Comitê da pós-graduação do IPEN.

Ainda com relação à produtividade das teses e aos artigos publicados, decidiu-se efetuar um exercício assinalando outro cenário temporal, para auxiliar a visualização, objetivando evitar vieses nos resultados; assim, foi realizado um corte em um período intermediário de 20 anos; observe-se que há diferença entre a média total e a média de artigos correlacionados, e estes dados também mostram tendência a crescimento.

A inexistência de trabalhos de análise da produtividade de artigos a partir das teses não permitiu comparações; no entanto, o Ministério de Ciência e Tecnologia (MCT) estabeleceu, a partir de 2002, um *Termo de Compromisso de Gestão* com todas as suas Unidades de Pesquisa (UPs) de administração direta do Ministério de Ciência e Tecnologia. Esse *Termo* se constitui em um instrumento de acompanhamento e avaliação anual de desempenho científico, tecnológico e gerencial por meio de indicadores e respectivas metas mutuamente pactuadas (BRASIL, 2007).

Dentro desse *Termo* solicitam-se indicadores de desempenho da Produção Científica. Um dos indicadores diz respeito ao número de “publicações vinculadas às dissertações e teses”, e tanto podem ser artigos em periódicos como em eventos nacionais e internacionais. Examinando-se os gráficos de uma das suas unidades, o Instituto de Pesquisas Espaciais (INPE), por exemplo, participante do *Termo de Gestão*, percebe-se que se trata de uma distribuição e contabilização numérica, não se observando uma análise mais detalhada ou correlacionada, como a que foi realizada pela técnica de coocorrência, e apresentada na FIG. 25 p. 126. Acredita-se que neste ponto é que reside a maior contribuição deste trabalho.

Certamente, os indicadores, constantes no Termo de Gestão do MCT, se constituem em uma fonte de referência que poderá ser utilizada após um tempo de consolidação dos dados para permitir um trabalho de análise comparativa de séries. Para ser utilizada, também exigirá um cuidadoso trabalho de estratificação, padronização e consistência dos dados.

Ainda com relação ao contexto global para a elaboração de indicadores de produção científica e com relação à ausência de trabalhos similares, a constatação da autora do trabalho é validada pela seguinte afirmação:

O conhecimento científico e técnico produzido no país não conta, até o presente momento, com estudos de institucionalização cognitiva e social do conhecimento produzido nas universidades e institutos de pesquisa. Não se dispõe, dessa forma, de estudos abrangentes sobre as temáticas de pesquisa dos programas de pós-graduação, as relações sociais estabelecidas (cooperação) e seu mapeamento. Semelhante situação é responsável, em larga medida, pelas dificuldades enfrentadas pelas universidades e institutos de pesquisa para avaliarem sua própria produção e produzir indicadores confiáveis. Comprometem-se, dessa forma, suas políticas de pesquisa, incluídas aí a definição de áreas ou projetos prioritários para destinação de recursos (KOBASHI, 2006).

6.4 O Currículo Lattes

Considera-se oportuno registrar a experiência sobre a utilização do Currículo Lattes. Em primeiro lugar, o trabalho de levantamento dos dados teve que ser realizado de forma manual (foram coletados os dados de 536 teses), uma vez que não há possibilidade de se obterem os dados institucionais de forma automática, embora a Direção do IPEN já houvesse, em ocasiões anteriores, solicitado uma autorização para acessar o Lattes institucional, sem sucesso.

Após a operação de coleta de dados manual foi possível constatar que apesar de o CL se constituir em importante fonte de informação pelo volume de dados que acumula sobre a comunidade científica brasileira, única no país e na América Latina,

apresenta inconsistências significativas que requerem cautela para aqueles que desejem utilizar estes dados em estudos bibliométricos, isto é, os dados recuperados do CL demandam um trabalho extra de formatação, consistência, complementação e padronização, esta inconsistência é mais notável, quanto mais se retroage na data de coleta dos dados.

Confirma-se e concorda-se com a constatação apresentada na tese de SILVA (2007), em que o autor registra que, por ser o CL um sistema aberto (na terminologia da Ciência da Informação, isto significa que o próprio autor realiza o autoarquivamento), apresenta um baixo nível de controle, no aspecto que se refere às representações linguísticas que são utilizadas no preenchimento dos Currículos, como: a) normalização gramatical; b) controle de homonímia; c) controle de sinonímia; d) opções de grafia. Embora esse formato adotado no desenvolvimento do Currículo Lattes apresente grande benefício no ganho econômico ao reduzir os investimentos na sua manutenção, perde em consistência nas informações disponíveis (SILVA, 2007).

6.5 As fontes de informação

No decorrer da elaboração do trabalho, esta autora teve a oportunidade de conhecer não só a literatura disponível sobre o uso das técnicas bibliométricas, mas também especialistas, sobre o assunto por meio da participação em eventos nacionais e internacionais (ISSI, 2009)³⁹, professores/pesquisadores empenhados no uso e, principalmente, no desenvolvimento de técnicas de mensuração avançada para serem aplicadas na área científica. É uníssona a opinião destes no que se refere ao problema e à necessidade de preparação dos dados, constatando-se que realmente é a etapa que consome mais tempo em um trabalho como este. É visível no país a ausência de fontes de informação diversificadas abrangentes (temática e temporalmente, como bases de dados nacionais, por exemplo de onde se possam obter dados para aplicar as técnicas bibliométricas.

³⁹ *Internacional Conference on Scientometrics and Informetrics*, realizada no Rio de Janeiro em 14-17 julho de 2009.

Após a realização deste trabalho pode-se perceber que a principal fonte utilizada para realização dos estudos bibliométricos e cientométricos ainda é o *Web of Science* e/ou *Scopus* (em menor grau); estas bases permitem a extração dos dados estruturados prontos para tabulação, análises, confecção de mapas, *clusters*, redes, e assim por diante.

No país, mantida a estrutura e continuidade da Biblioteca Brasileira de Dissertações e Teses (BDTD) do IBICT, esta se constitui em uma potencial fonte de dados nacional para se obterem dados e se realizarem estudos bibliométricos com as dissertações e teses.

No entanto, na opinião da autora e após a realização deste trabalho, é preciso ressaltar que o uso e a aplicação das técnicas bibliométricas como instrumentos meio produzem resultados extraídos de dados disponíveis e devem ser utilizados com ponderação; nada substitui o processo de interpretação e validação dos especialistas que conhecem a estrutura e a dinâmica das áreas estudadas, evitando assim o uso indiscriminado dos resultados.

Esses comentários são pertinentes e objetivam alertar e incentivar os profissionais que decidam realizar trabalhos similares. Nesse sentido, já há uma experiência bem-sucedida na Biblioteca Virtual da FAPESP, criada em 2003, com o objetivo de preservar e disseminar a memória institucional daquela organização. Esse sistema de informação já foi planejado e preparado para fornecer indicadores de produção científica a partir da sua base de dados, tendo sido publicados alguns trabalhos nesse assunto (KRZYZANOWSKI et al., 2008).

6.6 O potencial das bases de dados e a utilização da LC para estudos bibliométricos

Durante todo o tempo de existência e elaboração da PTC acreditava-se estar desenvolvendo um trabalho de qualidade; isso é verdadeiro se a PTC for utilizada como instrumento de busca, recuperação e armazenamento da Produção Científica

da Instituição. Até na questão da preservação da memória científica, constatou-se que, pelo menos com relação aos artigos produzidos, havia uma defasagem de 12% de itens.

Quando se verificou o preparo da base de dados para uso dos dados e a aplicação das técnicas de bibliometria, a situação não foi tão positiva; os dados, apesar de corretos, demandaram um trabalho razoável de padronização e correção. Vale lembrar que o processamento técnico dos documentos, mais especificamente a indexação utilizando o vocabulário controlado, é fruto do trabalho de vários profissionais, e que, por mais padronizados que os procedimentos estejam, há sempre um componente subjetivo; ainda, durante os últimos 40 anos, o sistema passou por pelo menos dois processos de migração para atualização de *software*, quando a movimentação dos dados é inevitável.

Conforme foi relatado no quadro teórico, as bases de dados foram criadas para busca e recuperação da informação; no entanto, demonstraram serem dispositivos multifuncionais com grande potencial de utilização para se obterem dados para elaboração de indicadores, principalmente as bases que abrigam a produção científica de uma Instituição. A qualidade, seja no processo de recuperação, seja no de produção de indicadores, está largamente associada ao desempenho das bases de dados; essa qualidade depende acentuadamente de processos rigorosos de tratamento da informação, da estrutura adotada para extração dos metadados, das políticas de indexação e dos vocabulários controlados (IGAMI et al., 2007).

O uso das dissertações e teses como *corpus* de estudo confirmou o potencial dessa literatura cinzenta como matéria-prima, apropriada para elaboração de diversos tipos de indicadores científicos. No caso deste trabalho foram obtidos somente dois tipos de indicadores, porém dependendo dos objetivos de outra pesquisa, é possível ampliar este número. Conforme foi comentado na literatura as dissertações e teses se configuravam em literatura cinzenta por várias características, mas principalmente por: a) não participarem do circuito comercial das editoras – na opinião da autora deste trabalho, com base na experiência com o

trabalho da produção científica institucional, esta característica deve se manter, uma vez que um trabalho acadêmico nesse formato é incompatível com o mercado das editoras – (para se tornar comercial deve-se adequar o suporte como por exemplo à forma de um livro); b) com relação à dificuldade de localização, esta característica foi eliminada pelo avanço da tecnologia da informação e a consolidação dos repositórios institucionais. Há dez anos, quando surgiram os primeiros repositórios de dissertações e teses, como o da biblioteca digital de dissertações e teses da Universidade de São Paulo (USP), da qual o IPEN participa, foi necessário um grande trabalho de convencimento para que os autores depositassem voluntariamente seus trabalhos. Atualmente, esse depósito foi incorporado ao processo de homologação das teses e se tornou um processo rotineiro. Diante dessa constatação, de livre acesso as teses sem intermediação das editoras, arrisca-se a afirmar que a denominação *literatura cinzenta* deve permanecer, e a disponibilidade de acesso tende a se consolidar, e conforme foi demonstrado neste trabalho, trata-se de matéria-prima apropriada para realizar estudos bibliométricos.

Com relação à técnica utilizada, faz-se oportuno comentar que o uso dessa metodologia, análise de coocorrência de palavras-chave, em um primeiro momento, restringe-se àqueles sistemas que utilizam sistematicamente, para indexação dos trabalhos, um vocabulário controlado, como por exemplo, a área nuclear ou a área médica. Nesse sentido, as áreas exatas e biológicas são terreno mais apropriado para se desenvolverem estudos semelhantes pela própria natureza do seu conteúdo e pela preferência por publicação de resultados em artigos de periódicos. É fato reconhecido que as áreas de humanas, por exemplo, culturalmente adotam outros canais para comunicação da sua produção científica (MEADOWS, 1999).

Por outro lado, para se estabelecer coocorrência e identificar similaridade entre itens, é possível utilizar também outros elementos, como autores, palavras naturais, (extraídas dos resumos ou dos títulos) ano de publicação, assuntos, índices classificatórios, e assim por diante; enfim, dados que estejam presentes de forma consistente em todos os itens de literatura que se quer analisar e correlacionar. Independentemente de qual for a opção, todos os itens irão demandar um processo de consistência exaustiva.

É recomendável também a utilização de softwares de tratamento da informação para se tornarem material apto para uso na elaboração de indicadores. Os softwares especializados são ferramentas úteis para analisar a consistência dos dados e ainda minimizam a incidência de erros e agilizam o manuseio dos dados, principalmente quando se trabalha com um volume muito grande destes.

As bases de dados bibliográficas, principalmente as de Produção Científica, constituem-se em matéria-prima de qualidade para a elaboração de indicadores, não importando o estágio de uso de tecnologia em que elas se encontram ou o tipo de suporte que armazenam. Incluem-se aqui os novíssimos repositórios, os quais adotaram a filosofia, a estrutura e o processamento técnicos bastante similares às bases de dados, agregando valor com o autoarquivamento e a filosofia do acesso aberto (*Open access*).

Seria aconselhável que os gestores de bases de dados adotassem políticas de gestão de qualidade das bases, que permitissem “olhar” os dados de forma sistemática e promover um programa de consistência contínua, isto é, que se criasse uma cultura de monitoração. Certamente é um trabalho que requer inversão de tempo e recursos, no entanto, vale a pena se for considerado que se trata de um dispositivo que permite dar materialidade ao ciclo da informação e conhecimento (ou o *continuum* do conhecimento) citado por TAYLOR (1996), no seu modelo de valor agregado (FIG. 1, p. 19 deste trabalho).

7 CONCLUSÃO

Nos objetivos deste trabalho foi registrada a intenção de se desenvolver uma pesquisa que pudesse responder a algumas questões sobre o potencial das dissertações e teses para se tornarem dados úteis para a elaboração de indicadores científicos, bem como sobre a multifuncionalidade das bases de dados, um dos mais conhecidos dispositivos da memória institucionalizada.

Com relação ao trabalho de elaboração de indicadores científicos, conclui-se que os resultados da análise longitudinal contribuíram para evidenciar a dinâmica e o perfil do trabalho acadêmico realizado no âmbito do Programa de PG da Instituição nestas três décadas de existência. No período estudado de 1977 a 2009, o desempenho numérico, temático e produtivo está mapeado e servirá de base para estudos futuros; em outras palavras, mapearam-se as tendências da institucionalização social da pesquisa científica realizada no Programa.

A análise temática temporal registrou a atividade científica dos diversos momentos da história da Instituição, positivos ou não tão positivos. Os dados permitiram correlacionar os resultados com os fatos históricos, políticos e econômicos.

As categorias temáticas agrupadas indicam que as dissertações e teses elaboradas na área nuclear se constituem em 51% de todas as outras áreas, no período 1977-2009. Isto demonstra que o Programa de Pós-graduação do IPEN não perdeu o foco, embora em determinados períodos, por decisão política, o ritmo deste tenha sido desacelerado. Já na apresentação dos dados distribuídos por áreas temáticas sem agrupamento, mostra um cenário onde se destaca a multidisciplinaridade dos trabalhos, principalmente em Ciência dos Materiais, Química, Aplicações das Técnicas Nucleares, Física e Reatores.

O estudo evidenciou as áreas temáticas que estiveram presentes na trajetória do Programa, aquelas consideradas “frentes de pesquisa”, áreas que mantiveram a mesma atuação e passaram incólumes às situações políticas adversas (Ciência dos Materiais) ou aquelas que “sofreram” a influência das mudanças políticas,

interrompendo sua trajetória, porém, reagindo posteriormente (Química e Reatores). O mapeamento das teses após o tratamento estatístico de análise de clusters permitiu reunir os trabalhos em cinco grupos, demonstrar a evolução na trajetória e visualizar tendências futuras de cada grupo. Finalizando o estudo foi possível perceber que a produção do conhecimento é um processo dinâmico, e os resultados obtidos nesta oportunidade se limitam a retratar um momento da produção científica do Programa de PG, os dados não devem ser tomadas como verdades inalteráveis, e sim como representações históricas obtidas após a aplicação de “filtros”

Conseguiu-se também demonstrar o perfil da produtividade das teses no que se relaciona à publicação de artigos em periódicos; observou-se de uma forma mais realística o volume de artigos que dão prosseguimento aos temas desenvolvidos nas teses, subsidiando novas pesquisas ou integrando-se ao mundo científico internacional.

Conclui-se que no universo estudado, 55,9% dos artigos publicados são fortemente correlacionados com a tese de origem, conforme foi demonstrado pela aplicação do método automático.

No período intermediário estudado, 1985 a 2004, observa-se um acréscimo na média de artigos publicados e correlacionados com as teses, alcançando no quinquênio de 2000 a 2004 uma média de 4,8 artigos por tese, deve ser olhada com cautela, trata-se de uma média quinquenal embora muito próxima da realidade e respeitadas as devidas ponderações, trata-se de uma média significativa e está próxima da política de publicações estabelecida pelo Programa de PG

Foi possível observar que o tempo para publicação dos artigos derivados das teses está sendo significativamente abreviado. Em 1992, os artigos correlacionados com as teses levavam em média entre 1,5 a 2,05 (ano) para serem publicados, observou-se que a partir do ano de 2004 houve uma queda acentuada neste tempo e, em 2009, os artigos já são publicados, em média, 1,63 (ano) antes da defesa da tese (embora esse dado tenha que ser visto com ressalvas); demonstrou-se também que os canais utilizados para divulgação dos artigos propiciam a visibilidade internacional dos trabalhos.

A técnica de coocorrência de palavras, aplicada e validada pela opinião dos autores das teses, mostrou-se adequada, uma vez que “filtra” melhor e evidencia a correlação entre os artigos produzidos pelas teses. A opção de estabelecer uma correlação mais forte a partir da coocorrência de três palavras-chave entre a tese e o artigo mostrou-se satisfatória, conforme foi demonstrado pelo cálculo de intervalo de confiança de 95%. Após ser submetida a testes em outras áreas ou com outros elementos coocorrentes, indica que tem potencial para se constituir em uma ferramenta adequada para ser utilizada com a finalidade de aprofundar a simples operação tradicional de enumeração de dados, sem os devidos critérios de ponderação.

Com relação à viabilidade de se utilizarem as bases de dados como fornecedoras de dados para a mensuração da produção científica, conclui-se que é possível diversificar o uso e explorar o potencial das BDs como instrumentos para desenvolverem análises bibliométricas. Demonstrou-se que a estrutura dos metadados e a qualidade dos registros armazenados nesses dispositivos são pré-requisito para se apresentar um bom trabalho, ou, em outras palavras, interferem na confiabilidade dos estudos. Observe-se o caso da PTC utilizada neste trabalho como campo de experimentação, certamente, a estrutura adotada e mantida desde a criação da base, a metodologia de trabalho, qualidade dos dados e o longo tempo de existência da base (37 anos) favoreceram a realização deste estudo.

7.1 Limitações do estudo e sugestões para futuros trabalhos

É importante comentar que neste trabalho foram analisadas somente as dissertações e teses produzidas pelo Programa da PG, porém, se trata de **uma** das modalidades de literatura existentes; o estudo não se estendeu para a análise dos artigos apresentados em eventos nacionais ou internacionais, constituindo-se em uma limitação do estudo; a PTC encerra um número expressivo (9,100 itens) desse tipo de literatura e certamente demandaria tempo maior para preparação dos dados. A situação é semelhante no que diz respeito aos demais itens de literatura que compõem a PC da Instituição.

A ausência de estudos semelhantes constitui-se também em uma limitação para o estudo, estes propiciariam uma comparação na abordagem utilizada. Para se obter uma visão mais ampla, também se sugerem estudos futuros repetindo as análises para confirmarem ou verificarem a “movimentação” das áreas, numérica e tematicamente, bem como o nível de produtividade das teses com relação à publicação de artigos.

Conforme foi comentado na Introdução deste trabalho, não foi intenção elaborar indicadores de citação; no entanto, na atual situação, com o volume de dados coletados e preparados para o campo de experimentação, vislumbra-se a oportunidade de se elaborarem futuros trabalhos como:

- identificação do índice de citação dos artigos correlacionados, o que indicaria o grau de relevância dos assuntos abordados e desenvolvidos nas teses;
- subdivisão das áreas temáticas gerais em subáreas, o que proporcionaria uma visão mais micro do cenário estudado;
- elaboração de clusters reunindo orientadores, teses e número de artigos, o que indicaria o grau de participação dos orientadores na sequência dos trabalhos durante e após tese;
- comparação da técnica de coocorrência, desta vez utilizando as palavras-chave fornecidas pelos próprios autores das teses, com as palavras-chave atribuídas pelo indexador para os artigos, o que reforçaria a validade da técnica.

APÊNDICE A – Exemplo do questionário enviado para os autores das teses

Prezado autor, XXXXX:

Estou desenvolvendo um estudo sobre a produtividade das teses, com relação à publicação de artigos, defendidas no Programa de PG do IPEN no período de 1977 até 2009, como parte do meu trabalho de Doutorado no IPEN, sob orientação do Dr. Bressiani. Para tanto, peço a sua colaboração no sentido de indicar se os seguintes artigos publicados têm correlação com a sua tese: ***Avaliação neutrônica de reator carregado com combustível metálico e refrigerado por chumbo***, defendida em 2000.

Para facilitar a resposta, utilize esta mesma mensagem, para tanto, clique primeiro em responder, depois assinale com um “X” entre os colchetes a correlação do artigo, clique em enviar, não é necessário salvar nada.

Esclareço que o estudo contempla os artigos publicados somente no período de 5 anos antes e 5 depois do ano de defesa da tese.

Agradeço a sua colaboração e me coloco à disposição.

Mery P. Zamudio Igami

Bibliotecária

IPEN – mery@ipen.br

Fone: 3133 9094

Artigo 1 de 2

SANTOS, A; NASCIMENTO, J.A. An Integral lead reactor concept for developing countries. **Nuclear. Technology**, v. 140, p. 1-22, 2002.

- 1- [X] o artigo publica os resultados obtidos durante ou após a elaboração da tese, isto é, tem uma alta correlação com a tese.
- 2- [] o artigo publica grande parte dos resultados da tese, porém, aplicados ou complementados com o resultado de outras pesquisas, isto é, tem uma correlação média com a tese.
- 3- [] o artigo publica resultados de outras pesquisas e só aproveita a experiência da tese, isto é, tem uma pequena correlação com a tese.
- 4- [] o artigo publica resultados de pesquisas sem nenhuma correlação com a tese.

Artigo 2 de 2

SANTOS, A; NASCIMENTO, J.A. An integral metallic-fueled and lead-cooled reactor concept for the 4th generation reactor. **J. Nuclear. Science & Technology** 2002. Supl. 2. p. 1081-1084.

- 1- [X] o artigo publica os resultados obtidos durante ou após a elaboração da tese, isto é, tem uma alta correlação com a tese.
- 2- [] o artigo publica grande parte dos resultados da tese, porém, aplicados ou complementados com o resultado de outras pesquisas, isto é, tem uma correlação média com a tese.
- 3- [] o artigo publica resultados de outras pesquisas e só aproveita a experiência da tese, isto é, tem uma pequena correlação com a tese.
- 4- [] o artigo publica resultados de pesquisas sem nenhuma correlação com a tese.

APÊNDICE B – Tabela matriz construída a partir das informações extraídas do Currículo Lattes e da PTC

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
S43	11132	CALVO, W.A.P.	Desenvolvimento do sistema de irradiação em um irradiador multipropósito de cobalto-60 tipo compacto	IRRADIATION PLANTS; IRRADIATION DEVICES; GAMMA RADIATION; COBALT 60; GAMMA SOURCES; MODULAR STRUCTURES; RADIATION SOURCES; INDUSTRIAL PLANTS.	2005	ANDRADE E SILVA, L.G.	Gamma-ray computed tomography scanners for applications in multiphase system columns	CALVO, W.A.P.; HAMADA, M. M. ; SPRENGER, F. E. ; VASQUEZ, P. A. S. ; RELA, P. R. ; MARTINS, J. F. T. ; PEREIRA, J. C. S. M. ; OMI, N. M. ; MESQUITA, C. H.	Nukleonika	2009	POROSITY; GAMMA RADIATION; COBALT 60; RADIATION SOURCES; PERFORMANCE; MATERIALS TESTING; RADIATION DETECTORS; COMPUTERIZED TOMOGRAPHY	14922	PI	WoS
S43	11132						Development of an irradiation system for a small size continuous run multipurpose gamma irradiator	CALVO, W.A.P.; RELA, P.R. ; NAPOLITANO, C. M. ; KODAMA, Y. ; OMI, N. M. ; COSTA, F. E. da ; ANDRADE E SILVA, L.G.	Nukleonika	2009	IRRADIATION PLANTS; IRRADIATION DEVICES; GAMMA RADIATION; COBALT 60; GAMMA SOURCES; MODULAR STRUCTURES; BRAZIL; RADIATION SOURCES; INDUSTRIAL PLANTS.	14929	PI	WoS
S43	11132						Desenvolvimento e implantação de um irradiador multipropósito de cobalto-60 tipo compacto	RELA, P.R; CALVO, W.A.P.; SPRINGER, F.E; OMI, N.M; COSTA, F.E; VIEIRA, J.M; ANDRADE E SILVA, L.G.	Revista Brasileira de Pesquisa e Desenvolvimento	2003	IRRADIATION PLANTS; COBALT 60; GAMMA RADIATION; IONIZING RADIATIONS; MODULAR STRUCTURES.	9450	PI	WoS
S43	11132						Activity measurements of sup(192)Ir solid sources using a well-type ionization chamber	GONÇALVES, J.A.C.; BOTELHO, S.; PASCHOLATI, P.R.; RIDENTI, M.A.; FRAGA, M.M.F.R.; CAMARA, J.R.; CALVO, W.A.P.; BUENO, C.C.	Nuclear Instruments and Methods in Physics Research. Section A	2007	IRIDIUM 192; TECHNETIUM 99; IONIZATION CHAMBERS; ACTIVITY LEVELS; DECAY; SEALED SOURCES.	12830	PI	WoS
S43	11132						A small size continuous run industrial gamma irradiator	CALVO, W.A.P.; RELA, P.R; SPRINGER, F.E; COSTA, F.E; OMI, N.M; ANDRADE E SILVA, L.G.	Radiation Physics and Chemistry	2004	CNEN; MODULAR STRUCTURES; COBALT 60; ELECTRON BEAMS; GAMMA RADIATION; GAMMA SOURCES; IRRADIATION DEVICES; MANUFACTURES; SHIELDING; TRANSPORT; VALIDATION.	9945	PI	WoS
S43	11132						Development of a scintillator detector system for gamma-ray scan measurements of industrial distillation columns	VASQUEZ, P.A; COSTA, F.E; CALVO, W.A.P.; HAMADA, M.M.	Nuclear Instruments and Methods in Physics Research. Section A	2005	CESIUM IODIDES; COBALT 60; DISTILLATION; FAILURES; GAMMA RADIATION; PERFORMANCE; PETROCHEMICAL PLANTS; SEALED SOURCES; SOLID SCINTILLATION DETECTORS.	10249	PI	WoS

Legenda: 1 – Categoria temática; 2 – Nº de IPEN-doc; 3 – Autor/tese; 4 – Título/tese; 5 – Palavras-chave/tese; 6 – Ano/tese; 7 – Orientador; 8 – Título/artigo; 9 – Autores/artigo; 10 – Título /revista; 11 – Ano/artigo; 12 – Palavras-chave /artigo; 13 – Nº IPEN-doc/artigo; 14 – Periódico internacional; 15 – Indexado na Web of Science (WoS)

Dados da tese

Dados dos artigos

REFERÊNCIAS BIBLIOGRÁFICAS

ABRÃO, A. ATALLA, L.T. **Determinação de urânio, tório, nióbio e tântalo em alguns minérios uraníferos de Araxá**. São Paulo: Instituto de Energia Atômica, 1962. (IEA-Pub- 53). Disponível em:< http://pintassilgo2.ipen.br/biblioteca/iea/IEA_PUB_42.pdf > Acesso em: jan.2010.

AGUIAR, P. Por uma imagem real da energia nuclear. **Brasil Nuclear**, v.14, nº 34, 2008.

ANDRADE, A. M. R. **A opção nuclear**. 50 anos rumo a autonomia. Rio de Janeiro: MAST, 1996.

ARANHA, F. O novo programa nuclear brasileiro. **Brasil Nuclear**, v.11, n.28, p. 16-20, 2005. Disponível em: < <http://www.aben.com.br/publicações> Acesso em: mar. 2009.

BASSECOULARD, E.; LELU, A.; ZITT, M. Mapping nanosciences by citation flows: a preliminary analysis. **Scientometrics**, v.70, n.2, p.859-880, 2007. < Disponível em: < <http://www.periodicos.CAPES.gov.br> > Acesso em: mar.2008.

BERNAL, J. D. **Historia social de la ciência**. La ciência en la historia. Barcelona: Península, 1973. v.1 p. 34.

BASSELAAR, P. VAN DEN; HEIMERIKS, G. Mapping research topics using word-reference co-occurrences: a method and an exploratory case study. **Scientometrics**, v.68, n.3, p.377-393, 2006. Disponível em: < <http://www.periodicos.CAPES.gov.br> > Acesso em: dez, 2009.

BOLFARINE, H.; BUSSAB, W. O. **Elementos de amostragem**. São Paulo: Edgar Blucher, 2005.

BORGES, M. M. **Os repositórios institucionais: viabilidade e interesse do “estudo geral”**. Coimbra: Universidade de Coimbra, 2008. (Repositório). Disponível em: > <http://www.sr.uc.pt> > Acesso em: dez. 2010.

BOURDIEU, P. **Os usos sociais da ciência: por uma sociologia clínica do campo científico**: Pierre Bourdieu. São Paulo: UNESP, 2004.

BRAAM, R. R. ; MOED, H. F ; Van RAAN, A. F. J. Mapping of science by combined co-citation and word analysis. **Journal of the American Society for Information Science, (JASIS)**, v.42, p. 233-266, 1991.

BRAGA, G. M. Informação, ciência e política científica: o pensamento de Derek de Solla Price. **Ciência da Informação**, v.3, n.2, p.155-177,1974.Disponível em: < <http://www.scielo.bireme.br> > Acesso em: mar. 2009.

BRASIL. Decreto nº 39872 de 31 de agosto de 1956. Cria o Instituto de Energia Atômica e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**. Poder Executivo, Brasília, DF 31, ago. 1956, Seção 1 p. 16588. Disponível em < <http://www2.camara.gov.br/legin/fed/decret/1950-1959/decreto-39872-31-agosto-1956-333144-norma-pe.html> > Acesso em dez. 2010.

BRASIL. Decreto nº 8691 de 28 de julho de 1993. Dispõe sobre o Plano de Carreiras para a área de Ciência e Tecnologia da Administração Federal Direta, das Autarquias e das Fundações Federais e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**. Poder Executivo, Brasília, DF, 28, jul. 1993, Seção1 p.10709. Disponível em: < . <http://www2.camara.gov.br/legin/fed/decret/1950-1959/decreto-39872-31-agosto-1956-333144-norma-pe.html> > Acesso em: dez. 2010

BRASIL. Decreto-Lei nº 2464 de 31 de agosto de 1988. Altera a denominação das Empresas Nucleares Brasileiras, S.A. – NUCLEBRAS, transfere bens de sua propriedade e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**. Poder Executivo, Brasília, DF 01 de setembro de 1988, Seção 1 p. 16793. Disponível em < <http://www2.camara.gov.br/legin/fed/decret/1950-1959/decreto-39872-31-agosto-1956-333144-norma-pe.html> > Acesso em: nov. 2010

BRASIL. Lei nº 5540 de 28 de novembro de 1968. Fixa normas de organização e funcionamento do ensino superior e sua articulação com a escola média e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**. Poder Executivo, Brasília, DF, 11 de nov., 1968, Seção 1 p.10369. Disponível em: < <http://www2.camara.gov.br/legin/fed/decret/1950-1959/decreto-39872-31-agosto-1956-333144-norma-pe.html> > Acesso em: out. 2010

BRASIL. Ministério de Ciência e Tecnologia. Indicadores: apresentação/histórico. Brasília, 2002. Disponível em < <http://www.mct.gov.br/index.php/content/> > Acesso em: abr. 2011.

BRASIL. Ministério de Ciência e Tecnologia. Secretaria Executiva. Subsecretaria de Coordenação das Unidades de Pesquisa. Termo de Compromisso de Gestão – TCG. Análise dos resultados obtidos. Brasília, 2007. < Disponível em : < <http://sigmct.mct.gov.br/index.php/content/view/388.html> > acesso em: jan. 2010.

BUNGE, M. **Ciência e desenvolvimento**. São Paulo: EDUSP, 1980 (O homem e a ciência, v.11)

BUSH, V. **Science the endless frontier**. A report to the President. Washington: Government Printing Office, 1945 Disponível em : < <http://www.nsf.gov/od/lpa/nsf50/vbush1945.htm> > Acesso em; jan.2008..

BUSTANI, J. M. Deve o Brasil ter um programa nuclear? **Pensar Brasil**, nº3 abr., p. 2021, 2005.

BUTTER, R. K.; NOYONS, E. C. M. MACKELLENBERGH, Van M.; LAINE T. Combined concepts maps and bibliometric maps: first explorations. **Scientometrics**, v. 66, n. 2, p.377-387, 2006. Disponível em: < <http://www.periodicos.CAPES.gov.br> > Acesso em: dez, 2009.

BUXTON, A. Database searching: history, current and future developments. In: CONGRESSO REGIONAL DE INFORMAÇÃO EM CIÊNCIAS DA SAÚDE 1st, São Paulo out. 14-16, 1992.

CABENA, P.; HADJNIAN, P.; STADLER, R.; VERHEES, J. ZANASI, A. **Discovering data mining**. New Jersey: Prentice Hall, 1998

CALLON, M. ; COURTIAL, J. P. ; LAVILLE, F. Co-word analysis as a tool for describing the network of interactions between basic and technological research. The case of polymer science. **Scientometrics**, v.22, n. 1, p. 155-205, 1991.

CANCHUMANI, R. M. L. A propósito da construção de indicadores de ciência e tecnologia (C&T) no Brasil: aspectos sociotécnicos. **Datagramazero**, v.11, n.3, jun. 2010. Disponível em : < <http://www.dgz.org.br/> > Acesso em: maio 2011.

CARVALHO, E. M. R. De. Grey literature and its contributions to knowledge society. In: 67th IFLA COUNCIL AND GENERAL CONFERENCE, 16-15, August, 2001.

CENDÓN, B. V. Bases de dados de informação para negócios. **Ciência da Informação**, v. 31, n.2, p.30-43, 2002. Disponível em : < www.bireme.scielo.br > Acesso em jul. 2010.

COURTIAL, J. P. A co-word analysis of scientometrics. **Scientometrics**, v.32, n.3, p. 251-260, 1994. Disponível em: < <http://www.periodicos.CAPES.gov.br> > Acesso em: jan. 2009.

DING, Y. ; CHOWDHURY, G. G.; FOO S. Bibliometric cartography of information retrieval research by using co-word analysis. **Information Processing & Management**, v. 37, n.6, p. 817-842, 2001. Disponível em: < <http://www.periodicos.CAPES.gov.br> > Acesso em: fev. 2009.

EGGHE, L. Expansion of the field Informetrics: origins and consequences. **Information Process & Management**, v.41, p.1311-1316, 2005.

EGGHE, L. ; ROUSSEAU, R. **Introduction to informetrics**: quantitative methods in library, documentation and information science . Amsterdam: Elsevier,1990.

ENERGIA NUCLEAR: custos de uma alternativa. **Comciência**. Revista eletrônica de jornalismo científico, nº12, ago., 2000. Disponível em: www.comciencia.br > Acesso em: fev. 2011

FERREIRA FILHO, A. L.; PASCHOA, A. O tratado de não proliferação nuclear (TNP). **VEJA**, maio 2010. Disponível em:< www.veja.abril.com.br> Acesso em: fev. 2011.

FINANCIADORA DE ESTUDOS E PROJETOS – FINEP. Termos e conceitos. Rio de Janeiro, 2002. Disponível em < www.finep.gov.br > Acesso em: julho 2007.

FUNARO, V. M. B. de O.; NORONHA, D. P. Literatura cinzenta: canais de distribuição e incidência nas bases de dados. In: POBLACION, D. A. ; WITTER, G.; SILVA, J. F. M. **Comunicação & produção científica**. Contexto, indicadores e avaliação. São Paulo: Angellara, 2006. Cap.8, p.215-234.

FREIRE-MAIA, N. A. **ciência por dentro**. 5.ed.Rio de Janeiro: Vozes, 1998.

FURTADO A. T. Novos arranjos produtivos, estado e gestão da pesquisa pública. **Ciência e Cultura**, v.57, n.1 p. 41-45, 2005.

GARDIN, J.C. La logique du plausible: essais d'epistemologie pratique. Paris: Maison, de Sciences de L'Homme, 1981 citado por SILVA, M. dos R. ; FUJITA, M. S. L. A prática da indexação: análise da evolução de tendências teóricas e metodológicas. **Transinformação**, v.16, n.2, p. 133-161, 2004. Disponível em: < <http://revistas.puc-campinas.edu.br/transinfo/viewissue.php?id=7> >Acesso em: abr. 2011.

GEISLER, E. **The metrics of science and technology**. Connecticut: Quorum, 2000.

GEWANDSZNAJDER, F. **O método nas ciências naturais**. São Paulo: Ática, 2010

GIL, A. C. **Métodos e técnicas de pesquisa social**. 3.ed. São Paulo: Atlas, 1991.

GLANZEL, W. A concise introduction to bibliometrics and its history. Leuven Belgica, Policy and Research Center for R&D indicators, Steunpunt O&O indicatoren, (2006) Disponível em < <http://www.steunpuntooi.be/bibliometrics.html> > Acesso em : abril, 2008.

GOLDFARB, A. M. A.; FERRAZ, M. H. M. Raízes históricas da difícil equação institucional da ciência no Brasil. **São Paulo em Perspectiva**, v.16, n. 3, 2002. Disponível em :< <http://www.scielo.bireme.br> > Acesso em jul. 2008.

GONÇALVES, A.; COSTA RAMOS, L. M. S. V.; CASTRO, R. C. F. Revistas científicas: características funções e critérios de qualidade. In: POBLACION, D. A. WITTER, G. P.; SILVA, J. F. M. da **Comunicação e produção científica**: contexto, indicadores, avaliação. São Paulo: Angellara, 2006. Cap. 6 p. 163-190.

GONSALEZ-SAEZ, S. **Aplicação de métodos bibliométricos e da co-word analysis na avaliação da literatura científica brasileira em ciências da saúde de 1990-2002**. 2005. Tese (Doutoramento) Faculdade de Saúde Pública, São Paulo. Disponível em: < <http://www.teses.usp.br> > Acesso em: jan. 2008.

GORDON, A. M. P. L. **Instituto de Pesquisas Energéticas e Nucleares (1956-2000). Um estudo de caso à luz da história da ciência, da tecnologia e da cultura brasileira**. 2003. Tese (Doutoramento) - Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, São Paulo. Disponível em: < <http://pintassilgo2.ipen.br/biblioteca/teses/11602.pdf> > Acesso em: jan. 2010.

GRACELLI, A.; CASTRO, C. de M. O desenvolvimento da pós-graduação no Brasil. **Ciência e Cultura** v.37, n.7 Supl. , p.188-201, 1985.

GRANGER, G. G. **Por um conhecimento filosófico**. Campinas: Papiro, 1989.

GUIMARÃES, J. A. A pesquisa médica e biomédica no Brasil. Comparações com o desempenho científico brasileiro e mundial. **Ciência e Saúde Coletiva**, v.9, n.2, p. 303-327, 2004. Disponível em : < www.bireme.scielo.br > Acesso em jun. 2010.

GUIMARÃES, J. A. Eixos do Plano Nacional de Pós-graduação 2011-2020. In : **C,T&I Brasil 2011-2015**. Desafios e perspectivas, contexto nacional e internacional. Fórum Nacional do CONCECTI/CONFAP, Fortaleza, nov. 2010. (Palestra).

GUIMARÃES, R. Pesquisa no Brasil. A reforma tardia. **São Paulo em perspectiva**, v.16, n.4, 2002 . Disponível em : < www.bireme.scielo.br > Acesso em jun. 2008.

HAYATI, D.; KARAMI, E. SLEE, B. Combining qualitative and quantitative methods in the measurement of rural poverty. **Social Indicators Research**, v.75, p.361-394, 2006. Disponível em:< <http://www.periodicos.CAPES.gov.br> > Acesso em : jan.. 2011.

HE, Q. Knowledge discovery through co-word analysis. **Library Trends**, v.48, n.3, p.133-159, 1999. Disponível em: < <http://www.periodicos.CAPES.gov.br> > Acesso em : dez. 2008

HOOD, W. ; WILSON, C. Informetric studies using databases. Opportunities and challenges. **Scientometrics** , v.58, n.3 p. 587-608, 2003. Disponível em : < <http://www.periodicos.CAPES.gov.br> > Acesso em : jan. 2008,

IGAMI, M. P. Z.; NOZAKI, P.; KOBASHI, N. Y. A gestão da qualidade em bases da produção científica. In: Congresso brasileiro de Biblioteconomia e Documentação, 22., 08 a 11 de julho, 2007, Brasília. **Anais...** São Paulo: FEBAB 2007. (XXII CBBB).

INGWERSEN, P.; WORMELL, I. Databases as analytical tools in research management: a case study. In: CRONIN, B. ; TUDOR- SILOVIC N. (Ed.) **The knowledge industries; levels of economic and social developments in the 1990s**. London: Aslib,1990, p. 205-216.

INGWERSEN, P. **Information retrieval interaction**. London: Taylor Graham, 1992. Disponível em: < <http://db.dk/pi/in>> Acesso em abr. 2011.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIENCIA E TECNOLOGIA. Biblioteca digital brasileira de teses e dissertações. Brasília. Disponível em: < www.ibict.br > Acesso em: abr 2010.

INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES. Ata da Comissão de Pós-graduação. São Paulo: IPEN/CNEN-SP, 2001 (Documento interno)

INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES. Ata da Comissão de Pós-graduação. São Paulo: IPEN/CNEN-SP, 2005 (Documento interno)

INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES. Relatório de gestão, 2007-2010. São Paulo: IPEN/CNEN-SP, 2007. (Relatório interno)

INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES. Plano diretor, 2002. São Paulo: Divisão de Planejamento e Inovação tecnológica, IPEN/CNEN-SP, 2002. . (Relatório interno)

INTERNATIONAL ATOMIC ENERGY AGENCY. Subject categories and scope descriptions. Vienna: IAEA, 2002 (ETDE and INIS joint reference series,2)

INTERNATIONAL NUCLEAR INFORMATION SYSTEM. Opening of the access to the INIS Database on the Internet. **Newsletter- Nuclear Information and Knowledge**, jun.2009 Disponível em :< [http:// www.iaea.org/INIS/Products](http://www.iaea.org/INIS/Products) > Acesso em :dez. 2010

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION- ISO. Information et documentation. Indicateurs de performance des bibliothèques, Genebra: ISO,1998.(ISO 11620:1998(F). Disponível em : < <http://www.iso.org> > Acesso em : jan. 2008.

JARAMILLO, H.; LUGONES, G.; SALAZAR, M. **Manual de Bogotá**: normalización de indicadores de innovación tecnológica en America latina y el Caribe. Bogotá: RIYCT, 2001.

JONES, B. NAGIN, D. ROEDER, K. A SAS procedure based on mixture models for estimating developmental trajectories. **Sociological Method Research**, v.29, p. 374-393, 2001.

KENT, A. **Manual de recuperação mecânica da informação**. São Paulo: Polígono, 1972

KOBASHI, N. Y. ***Temas e linhas de pesquisa em Ciência da Informação***. Projeto de pesquisa desenvolvido na Universidade de São Paulo. São Paulo: Universidade de São Paulo, 2004.

KOBASHI, N. Y. & SANTOS, R. N. M. Institucionalização da pesquisa científica no Brasil: cartográfica temática e de redes sociais por meio de técnicas bibliométricas. ***Transinformação***, Campinas, v.18, n.1, p. 27-6, 2006. Disponível em: < <http://revistas.puc-campinas.edu.br/transinfo/viweissue.php?id=7> > Acesso em: abr. 2010.

KOBASHI, N. Y. ; SANTOS, R. N. M. Arqueologia do trabalho imaterial: uma aplicação bibliométrica à análise de dissertações e teses. ***Encontros de Biblioteconomia: Revista Eletrônica de Biblioteconomia e Ciência da Informação***, n. esp., 1º sem. P. 106-115, 2008.

KRZYŻANOWSKI, R. F.; IMPERATRIZ, M. I. de M. ; IGAMI, M. P. Z. ; TARUHN, R. A BV/FAPESP como subsídio para a produção de indicadores: artigos científicos em Ciências da Saúde e Biológicas. In: CONGRESSO REGIONAL DE LA INFORMACIÓN EN CIÊNCIAS DE LA SALUD – CRICS, 2008, Rio de Janeiro. São Paulo : Fapesp, 2008. (Poster).

KURAMOTO, H. Acesso livre: maximizando a visibilidade das pesquisas e dos pesquisadores. In: 3º SEMINÁRIO EM CIÊNCIAS DA INFORMAÇÃO (III SECIN), Londrina , Paraná , nov. 2009. (palestra). Disponível em: <http://kuramoto.blog.br> > Acesso em dez. 2010.

KURAMOTO, R. e APPOLONI, C. R2002. Uma breve história da política nuclear brasileira. ***Cadernos Brasileiros de Ensino de Física***, v.19, n.3, p. 379-392,

LA FUENTE, L. La movilización de la ciência. ***Quark***, n.28/29, abr.-set. 2003. Disponível em : < <http://www.imim.es/quark/28-29/028012.htm>. > Acesso em: abr. 2006.

LANCASTER, F. W. **Indexação e resumos**: teoria e pratica. Brasília: Briquet de Lemos, 1993.

LEVY, P. **A conexão planetária: o mercado, o ciberespaço, a consciência**. São Paulo: Editora 34 , 2001..

LIBERAL, C. G. Indicadores de ciência e tecnologia: conceitos e elementos históricos. **Ciência e Opinião**, v. 2, n.1/2, p. 121-141, 2005. Disponível em: < <http://cienciaeopinio.up.edu.br> > Acesso em: maio 2011.

LIMA, F.W; ABRAO, A. **Produção de compostos de urânio atômicamente puros no IEA**. São Paulo: Instituto de Energia Atômica, 1961. (IEA-Pub- 42). Disponível em:< http://pintassilgo2.ipen.br/biblioteca/iea/IEA_PUB_42.pdf > Acesso em: jan.2010.

LOPES, I. L. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. **Ciência da Informação**, v.31, n. 1, p.41-52, 2002. Disponível em: < <http://www.scielo.bireme.br> > acesso em: dez. 2008.

LOUREIRO, M. D. **Energia nuclear**. Rio de Janeiro: Bloch, 1980.

LOUREIRO-ALVES, M. de F. **Conhecendo um campo de estudo: aspectos da institucionalização cognitiva e social da ciência da informação**. 2010. Tese (Doutoramento). Escola de Comunicações e Artes, Universidade de São Paulo. Disponível em : < <http://www.teses.usp.br> >Acesso em: jan. 2011.

LOURENÇO, C. A. Automação em bibliotecas: análise da produção via biblioinfo (1986-1994). In: Witter, G. P.(Org) **Produção científica**. Campinas: Átomo, 1997. p. 25.

MACHADO, G. W. **A transformação do conhecimento em produtos e serviços: a literatura cinzenta e as teses do PPGE**. 2007. Tese (Doutoramento). Florianópolis, Universidade Federal de Santa Catarina, SC.

MACIAS-CHAPULA, C. A. O papel da informetria e da cientometria e sua perspectiva nacional e internacional. **Ciência da Informação**. v. 27, n.2, 1998. Disponível em: < <http://www.scielo.bireme.br> > acesso em: dez. 2004.

MARQUES, P. Q. **Sofismas nucleares**. Sao Paulo: Hucitec, 1992.

MASON, M. Y. **Grey literature: history, definition, acquisition, and cataloguing**. Research paper. s. d. Disponível em > <http://www.moyak.com> > Acesso em : dez. 2009.

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

MENEGHINI, R. A avaliação da produção científica e o projeto SciELO. **Ciência da Informação, Brasília**, v. 27, n. 2, p. 219-220, maio/ago. 1998. Disponível em: < <http://www.scielo.bireme.br> > acesso em: dez. 2008.

MERTON, R. K. **The sociology of science**: an episodic memoir. Carbondale: Southern Illinois University, 1979.

MERTON R. K. **Sociología de la ciência**: investigaciones teóricas y empíricas. Madrid: Alianza, 1973.

MINAYO, M.C. de S ; SANCHES, O. Quantitativo-qualitativo: oposição ou complementaridade? **Cadernos de Saúde Pública**, v.9, n.3,p. 239-262, 1993. Disponível em: < <http://www.scielo.bireme.br> > acesso em: dez. 2010.

MOED, H. F. The use of databases for bibliometric analysis In: EGGHE, L.; ROUSSEAU, R. (Ed.) **Informetrics 87/88**. Amsterdam: Elsevier, 1988. p. 15-28.

MORAIS, R. de **Filosofia da ciência e da tecnologia**. 5.ed. São Paulo: Papyrus, 1988.

MORAVCSIK, M. J. Some contextual problems of science indicators. In: VAN RAAN, A. F.J. (Ed.) **Handbook of quantitative studies of science and technology**. Washington: Elsevier, 1988. Cap.1, p.11-30.

MUELLER, S. P. M. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da Informação** v.35, n.2, p. 27-38, 2006. Disponível em : < www.scielo.bireme.br > Acesso em : nov. 2008.

NAGIN, D. Analyzing developmental trajectories: a semi-parametric group based approach. **Psychological Methods**, v.4, p.139-177, 1999.

NASCIMENTO, A. C. S. **Mapeamento temático das teses defendidas nos programas de Pós-graduação em Educação Física no Brasil (1994-2008)**. 2010. Tese (Doutoramento) Escola de Comunicações e Artes – ECA, Universidade de São Paulo. Disponível em:< www.teses.usp.br > Acesso em: jan.2011.

NEFF, M. W.; CORLEY, E. A. 35 years and 160,000 articles: a bibliometric exploration of the evolution of ecology. **Scientometrics** v.80, n.3, p. 657-682, 2009. Disponível em: < <http://www.periodicos.CAPES.gov.br> > Acesso em: julho, 2009.

NEUFELD, M. L.; CORNOG, M. Database history: from dinosaurs to compact discs. **Journal of the American Society for Information Science**, v.37, n.4, p. 183-190, 1986. Disponível em: < <http://www.periodicos.CAPES.gov.br> > Acesso em: jan. 2009.

NORONHA, D. P. MARICATO, J. de M. Estudos métricos da informação: primeiras aproximações. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, n. especial, p.116-128, 2008. Disponível em: < <http://www.periodicos.ufsc.br> > Acesso em: abr. 2011.

NOYONS, C. M. E. **Bibliometric mapping as a science policy and research management tool**. 1999. Thesis (Ph.D). Leiden, DSWO Press, 1999. Disponível em:< <http://www.cwts.nl> > Acesso em: fev. 2006.

NOYONS, C. M. E. Bibliometric mapping of science in a science policy context. **Scientometrics**, v.50, n.1, p.83-98, 2001. < Disponível em: < <http://www.periodicos.CAPES.gov.br> > Acesso em: jan. 2008.

NOYONS E. C. M.; VAN RAAN, A. F. J. Science mapping from publications. An example in mathematics and computer science. In: MEIJ, J. ***Dealing with the data flood***. Mining data text and multimedia. The Hague: STT/Beweton, 2002. p.64-72

OFFICE OF SCIENTIFIC AND TECHNICAL INFORMATION. OSTI history. s.d. Disponível em : < <http://www.osti.gov>. > Acesso em: julho,2010.

ORGANISATION DE COOPERATION ET DE DEVELOPEMENT ECONOMIQUES – OECD. Glossary of statistical terms. Genebra: OECD, 2001 Disponível em: < <http://cs3-hq.oecd.org> > Acesso em: julho 2007.

ORGANIZAÇÃO MUNDIAL DE METEOROLOGIA. Mudança do Clima 2007: a base das ciências físicas. Sumário para os Formuladores de Políticas. Painel Intergovernamental sobre mudança do clima. Geneva, Suíça: OMM, 2007. Disponível em: < www.ipcc.ch > Acesso em: jan. 2011

PALIS JUNIOR, J. Um olhar sobre a ciência brasileira e sua presença internacional. ***Parcerias Estratégicas***, v.15, n. 31, p. 74-113, 2010. (Edição Especial) Disponível em :< www.cgee.org.br > Acesso em abr. 2011.

PEJSOVA, P. (Ed.) ***Grey literature repositories***. Praga: Radim Bacuvcik, 2010. Disponível em : < <http://www.verbum.name> > Acesso em: jan. 2011

PENA, S. D. J. O estado da ciência no Brasil: como dar um salto de qualidade? ***Parcerias Estratégicas***, v.15, n. 31, p. 116-127, 2010. (Edição Especial) Disponível em :< www.cgee.org.br > Acesso em abr. 2011.

PEREIRA, M. DE N. F.; RIBEIRO, C. J. S.; TRACTENBERG, L.; MEDEIROS, P. L. Bases de dados na economia do conhecimento: a questão da qualidade. ***Ciência da Informação***, v.28, n.2, p.215-223, 1999. Disponível em: < <http://www.scielo.org> > Acesso em: abr. 2010.

PINHEIRO, L. V. R.; SILVA, G. S. Cartografia histórica e conceitual da bibliometria/informetria no Brasil. In: CONFERÊNCIA Ibero-Americana de publicações eletrônicas no contexto da comunicação científica, 2. 17-21 nov., Rio de Janeiro 2008.

p.1-24 Disponível em < <http://cipecc2008.ibict.br/index.php/CIPECC2008/cipecc2008> > Acesso em set 2009..

POBLACION, D. A. Literatura cinzenta ou não convencional: um desafio a ser enfrentado. **Ciencia da Informação**, v.21, n.3, p. 243-246, 1992 Disponível em: < <http://www.scielo.bireme.org> > Acesso em: ago. 2010

POBLACION, D. A. ; OLIVEIRA, M. de *Input e output*: insumos para o desenvolvimento da pesquisa. In: POBLACION, D. A.; WITTER, G. P. ; SILVA, J. F. M. **Comunicação & produção científica**: contexto, indicadores e avaliação. São Paulo, Angellara, 2006. p. 57-80

PRITCHARD, A. Statistical bibliography or bibliometrics? **Journal of Documentation** , v.25, n.4, p.348-349, 1969.

PRONUCLEAR. Programa de recursos humanos para o setor nuclear. Dez anos de atividades 1977-1986. São Paulo, [s.d]. (Relatório interno).

RABELO, S. L. Proposta para a discussão da reestruturação do Programa Nuclear Brasileiro.Parte 2. **Jornal da Ciência**, v.690, mar. 2007.

RIP, A Mapping of science: possibilities and limitations. In: VAN RAAN, A. (Ed.) **Handbook of quantitative studies of science**. North Holland: Elsevier, 1988. Cap. 9 p. 253

ROCHA, R. R. Sistemas de recuperação de informações e mecanismos de busca na Web: panorama atual e tendências. **Perspectivas em Ciência da Informação**, v.11, n.2, p.161-173, 2006. Disponível em: < <http://www.scielo.bireme.org> > Acesso em: ago. 2009

ROKAYA, M. ; ATLAM, E.; FUKETA, M.; DORJI, T. C.; AOE, J. Ranking of field associations terms using co-word analysis. **Information Processing and Management**, v.44 p. 738-755, 2008. Disponível em: < <http://www.periodicos.CAPES.gov.br> > Acesso em: fev. 2009.

SANCHO, R. Indicadores bibliométricos utilizados en la evaluación de la ciencia y la tecnología. Revisión bibliográfica. **Revista Española de Documentación Científica**, v.13, n.3/4, p. 842-865, 1990. Disponível em < <http://redc.revistas.csic.es/index.php/redc> > Acesso em: ago.2009.

SANZ CASADO, E. ; MARTIN MORENO, C. Técnicas bibliométricas aplicadas a los estudios de usuarios. **Revista General de Información y Documentación**, v.7, n. 2, 1997. Disponível em: <http://revistas.ucm.es> . Acesso em nov.2008.

SCHWARTZMAN, S. Nota sobre a transição necessária da pós-graduação brasileira. Texto preparado como subsidio para a Comissão responsável pela elaboração do Plano Nacional de pós-graduação (PNPG) relativo ao período de 2011-2020. Rio de Janeiro, 2010. Disponível em : <http://www.schwartzman.org.br/sitesimon/> > Acesso em maio, 2011.

SCHMIEDECKE, W. C.; PORTO, P. A. PRONUCLEAR (1976-1986) e a formação de recursos humanos para a área de energia nuclear no Brasil. **Circumscribere**, v.4 p. 29-50, 2008

SCHWARTZMAN, S. A ciência da ciência. **Ciência Hoje** v. 2, n. 11 p. 54-50, 1984.

SCIENTIFIC ELECTRONIC LIBRARY ON LINE. Sobre Scielo. Disponível em<: www.scielo.bireme.br >Acesso em: abr. 2011.

SILVA, F. M. E. **Organização da informação em sistemas eletrônicos abertos de informação científica e tecnológica**: análise da plataforma Lattes. 2007. Tese (Doutoramento). Escola de Comunicações e Artes, Universidade de São Paulo. Disponível em : < <http://www.teses.usp.br> >Acesso em: jan. 2011.

SILVA, M. dos R. ; FUJITA, M. S. L. . A prática da indexação: análise da evolução de tendências teóricas e metodológicas. **Transinformação**, v.16, n.2, p. 133-161, 2004. Disponível em: < <http://www.scielo.bireme.org> > Acesso em: abr. 2011.

SPINAK, E. Indicadores cienciométricos. **Ciência da Informação**, v.27, n.2, p.15-23, 1998. Disponível em: < <http://www.scielo.bireme.org> > Acesso em: ago. 2007.

SPINAK, E. L. De La bibliometría tradicional a la cienciometría en la Web: son válidos los mismos paradigmas? Reflexiones primarias. In: CONFERÊNCIA IBEROAMERICANA DE PUBLICAÇÕES ELETRÔNICAS NO CONTEXTO DA COMUNICAÇÃO CIENTÍFICA, 17-21 nov., Rio de Janeiro, 2008. Disponível em <http://cipecc2008.ibict.br/index.php/CIPECC2008/cipecc2008> > Acesso em set 2009

STEFANIAK, B. Use of bibliographic databases for scientometrics studies. *Scientometrics*, v.12, p. 149-161, 1987.

SUAIDEN, E. Dimensão e perspectivas sociais do acesso livre à informação. *Ciência da Informação*, v. 35 n. 2, Editorial, 2006. Disponível em: <http://www.scielo.bireme.org> > Acesso em: nov.2009.

SVENONIUS, E. *Intellectual foundation of information organization*. Cambridge: MIT Press, 2001

SVENONIUS, E. Unanswered questions in the design of controlled vocabularies. *Journal of the American Society for Information Science*, v.37, n.5, p.331-340, 1986. Disponível em: < <http://www.periodicos.CAPES.gov.br> > Acesso em: maio. 2009.

TAGUE-SUTCLIFFE, J. An introduction to informetrics. *Information Processing & Management*. v.28, n. 1, p. 1-3, 1992.

TAMMARO, A. M. SALARELLI, A. *A biblioteca digital*. Brasília: Briquet de Lemos, 2006.

TAYLOR, R. *The value added information system*. Washington, Ablex, 1996.

TENOPIR, C. Reflections on two decades. *Library Journal*, v. 28, n.2, p.36, 2003. Disponível em: < <http://www.periodicos.CAPES.gov.br> > Acesso em: fev. 2009.

TENOPIR, C. Time to redefine databases. *Library Journal*, v.126, n.2, p.40-42, 2001. Disponível em: < <http://www.periodicos.CAPES.gov.br> > Acesso em: fev. 2009.

TRZESNIAK, P. Indicadores quantitativos: reflexões que antecedem seu estabelecimento. **Ciência da Informação**, v.27, n.2, , pp. nd-nd. 1998. Disponível em : <http://www.scielo.bireme.br> > Acesso em jan. 2007.

UNITED NATIONS, EDUCATIONAL, SCIENTIFIC AND CULTURAAL ORGANIZATION. Declaration on science and use of the scientific knowledge. In: WORLD CONFERENCE ON SCIENCE, 26 june-01 July, Budapest, Hungary, 1999. (Submitted to the 30th General Conference for adoption

VALLEJO RUIZ, M. V. **Estudio longitudinal de la producción española de tesis doctorales en educación matemática (1975-2002)**. 2005.Tesis (Doctoral). Universidad de Granada, España.

VAN RAAN, A F. J. Measurement of central aspects of scientific research performance. Interdisciplinary structure. **Measurement : Interdisciplinary Research and Perspectives**, v.3, n.1, p.1-19, 2005.

VAN RAAN, A F. J. The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments. **Technology assessment- Theory and Practice**, v.12, n.1, p.20-29, 2003.

VAN RAAN, A. F. J. Advanced bibliometric methods to assess research performance and scientific development: basic principles and recent practical applications. **Research Evaluation**, v.3, n. 3, p. 151-166, 1993.

VAN RAAN, A. J. F. VAN LEUWEN, T. N. Identifying the fields for mapping RTD excellence in life sciencies – a first approach. Report to the European Commission under contract COPO-CT-2001-00001, 2001.

VELHO, L. Indicadores de C&T no Brasil: antecedentes e estratégia. In : Cuarto Taller Iberoamericano de indicadores de Ciência y Tecnología, 12-14 de julio, México, 1999.

VINKLER, P. An attempt of surveying and classifying bibliometric indicators for scientometric purposes. **Scientometrics**, v.13, n. 5-6, p. 56-67, 1988.

VIOTTI, E. B. Doutores 2010: estudos da demografia da base técnico-científica brasileira. In: DOUTORES 2010: estudos da demografia da base técnico-científica brasileira. Brasília: CGEE, 2010. Cap. 1 p.19. Disponível em: <http://www.cgee.org.br> > Acesso em: maio.2011.

VOOS, H. Lotka and information science. *Journal of the American Society of Information Science*, v.25, p. 270-272, 1974. Disponível em: < <http://www.periodicos.CAPES.gov.br> > Acesso em: fev. 2010.

WHITLEY, R. Cognitive and social institutionalization of scientific specialties and research areas. In: WHITLEY, R. (Ed.) *Social processes of scientific development*. London: Routledge, and Kogan, 1974. p.69-75.

WHITTAKER, J. Creativity and conformity in science: titles, keywords and co-word analysis. *Social Studies in Science*, v.19 p. 473-496, 1989.

WILIAMS, E. M. The state of databases today: 2002. In: FAERBER, M. **Gale directory of data bases**. Detroit: Gale Group, 2002. vol.1 (Preface)

WITTER, G. P. (Org.) *Produção científica*. Campinas: Átomo, 1997.

WORMELL, I. Informetria: explorando bases de dados como instrumentos de análise. *Ciência da Informação*, v.27, n.2, p.210-216, 1998. Disponível em : < <http://www.scielo.bireme.br> > Acesso em: abr. 2009

XAVIER, R. S. A retomada do programa nuclear brasileiro e o gargalo organizacional. *Jornal da Ciência*, 30 de abril, 2009. Disponível em: < <http://www.jornaldaciencia.org.br/> > Acesso em: maio 2009.

YOON, B. LEE, S. LEE, G. Development and application of a key word-map based knowledge map for effective R&D planning. *Scientometrics*, v.85, p.803-820, 2010. Disponível em: > <http://www.periodicos.CAPES.gov.br> > Acesso em : jan.2011.

YUNTA, L. R. Bases de datos documentales: estructura y uso In: MALDONADO, A. (Coord.) *La información especializada en Internet*. Madrid: CINDOC, 2001 p.1 – 21.

YUNTA, L. R. Evaluacion e indicadores de calidad em bases de datos. **Revista Española de Documentación Científica**, v.21, n.1 p. 9-20, 1998. Disponível em: [http://redc.revistas.csic.es/index.php/redc >](http://redc.revistas.csic.es/index.php/redc) Acesso em: maio 2010

ZIMAN, J. M. **O conhecimento público**. Belo Horizonte: Itatiaia, 1979.