

Análise cienciométrica de produção científica por meio de Dissertações e teses: uma experiência brasileira

Raimundo Nonato Macedo dos SANTOS
Departamento de Ciência da Informação
Universidade Federal de Santa Catarina
raimundo@cin.ufsc.br

Nair Yumiko KOBASHI
Escola de Comunicações e Artes
Universidade de São Paulo – www.usp.br
nykobash@usp.br

José Carlos Bressiani
Instituto de Pesquisas Energéticas e Nucleares, IPEN – CNEN/SP
jbressia@ipen.br

Mery P. Zamudio Igami
Instituto de Pesquisas Energéticas e Nucleares, IPEN – CNEN/SP
mery@ipen.br

Resumo

Estudo de produção científica para conhecer aspectos da institucionalização cognitiva da atividade de pesquisa no Brasil, por meio da análise de dissertações e teses. A produção de indicadores, em moldes tradicionais, baseia-se na análise de artigos publicados em revistas de prestígio internacional. A utilização desses dados é recorrentemente criticada. Com efeito, não há relação direta entre a atividade intelectual produtiva e a publicação em periódicos internacionais, fato que parece ser aplicável à maioria dos países. Presume-se que há um trabalho intelectual altamente produtivo que antecede a publicação de artigos. É necessário, portanto, identificar as variáveis inerentes a esse processo, que sejam legítimas para descrever a atividade coletiva de pesquisa. Estas constatações solicitam a formulação de hipóteses alternativas de trabalho, adequadas para os contextos em que se desenvolve a atividade científica. Em nosso caso, partimos do pressuposto de que as dissertações e teses podem revelar aspectos importantes da atividade científica do país. Sustenta tal hipótese o fato de serem as dissertações e teses produzidas e avaliadas em instâncias acadêmicas altamente organizadas e controladas. No caso brasileiro, as dissertações e teses ganham maior significado porque são produzidas no espaço principal de pesquisa científica do país, que são os programas de pós-graduação. A metodologia de pesquisa utilizada combina métodos bibliométricos avançados e softwares de construção de mapas da ciência, que permitem visualizar temas de pesquisa. A pesquisa aplica os métodos bibliométricos e apresenta como solução, para os problemas encontrados, os procedimentos propostos pela área da Organização e Representação da Informação. Os resultados obtidos estão exemplificados em estudos realizados em duas áreas do conhecimento distintas de Energia Nuclear e Ciência da Informação. Os dados empíricos da pesquisa foram obtidos nas bases de dissertações e teses da CAPES e IPEN/SP. Além de indicadores sobre as temáticas de pesquisa, que são úteis para subsidiar políticas científicas, foram obtidos outros resultados importantes: a capacitação da equipe na elaboração de produção de indicadores e a sinalização de que os métodos bibliométricos são apropriados e necessários para a gestão de qualidade de bases de dados bibliográficas que registrem a Produção Científica.

Palavras-chave: Estudos de produção científica; Bibliometria; Mapas da ciência; Cienciométrica; Energia Nuclear; Ciência da Informação; Bases de dados bibliográficos.

Abstract

This scientific production study, intends to know some cognitive institutional aspects of the research activity in Brazil, using the dissertations and thesis as a research *corpus*. Traditional scientific output activities indicators are based on the analysis of the articles or papers published in the international journals. This practice is recurrently object of critics. As a matter of fact it seems not to exist a direct relation between the productive intellectual activity and the publishing in the international journals, fact which apparently, is true in the most of the countries. It is well known that there is an intense intellectual effort before to publish an article, therefore it is necessary to identify the inherent variables inserted in this research process to describe the collective research activity. Such kind of evidence requires formulation of alternative hypothesis of working, suitable to the context where the research activity is performed. In this paper it was assumed that a study including dissertations and thesis could reveal some interesting aspects of the national research activity, indeed, this kind of literature is produced in an appropriated scientific research *loci* as they are the Master and Doctoral Programs which are regulated and periodically assessed by a renamed academic national institution. The applied methodology combines advanced bibliometric tools and the use of softwares for the science maps construction which allow a general view of the thematic research fronts. The study proposes as a solution, for verifying inconsistencies, the procedures from Organization of Knowledge area. The results obtained are represented in studies performed in two distinct areas: Nuclear energy and Information Science. Empiric data were obtained from dissertations and thesis databases from CAPES and IPEN/SP respectively. Besides the thematic indicators, useful for definition of a scientific politics, it was obtained some other significant results as: the constitution of a multidisciplinary staff to develop studies of scientific activity, empowerment of information professionals for the production of databases, and the evidence that bibliometric tools are suitable and necessary for a quality management of the scientific production databases.

Key-words: Scientific activity studies, Scientific production, Bibliometrics; Science maps, Scientometrics; Nuclear Energy; Information Science; Bibliographic databases.

1. Introdução

Os indicadores da atividade científica, em sua maioria, decorrem da análise de artigos publicados em revistas de prestígio internacional. Essa metodologia é, porém, recorrentemente criticada, pois, há um trabalho intelectual altamente produtivo que antecede, substitui ou é concomitante à publicação de artigos. Com efeito, cada comunidade científica tem modos próprios de dar a conhecer a sua produção.

Pode-se afirmar, portanto, que não há relação direta entre a atividade intelectual produtiva e a publicação de artigos em periódicos internacionais, fato compartilhado pela maioria dos países. Nessa medida, é necessário identificar variáveis legítimas e adequadas para retratar a atividade coletiva de pesquisa de cada país. Estas constatações solicitam a formulação de hipóteses alternativas de trabalho, apropriadas para cada contexto de produção e comunicação do saber.

As abordagens bibliométricas e cienciométricas, modos específicos de mensurar a ciência e produzir indicadores, estão, segundo Macias-Chapula (1998), no centro do debate sobre os avanços da ciência, da tecnologia e do progresso econômico e social de cada país, sendo questionáveis as revisões ou reformulações de políticas científicas construídas ao largo dos indicadores de produção científica.

A construção desses indicadores é complexa já que deve ser considerado um conjunto heterogêneo de atividades, múltiplos agentes e instituições. Nessa medida, as especificidades de cada país indicam ser necessário redefinir pressupostos antes de empreender a tarefa de produzir indicadores. Conhecer em profundidade as características e hábitos de cada setor para elaborar hipóteses de trabalho plausíveis e, ao mesmo tempo, utilizar metodologias de coleta e análise que produzam resultados confiáveis, são o primeiro passo. No caso brasileiro, o desafio é maior, em face

da precariedade, ou mesmo ausência, de fontes de informação com a abrangência e a confiabilidade necessárias para a obtenção de resultados pertinentes. Portanto, antecede a produção de indicadores a coleta sistemática de dados e construção de repositórios *ad hoc* (Kobashi & Santos, 2006a).

No trabalho, ora apresentado, parte-se do pressuposto de que as dissertações e teses podem revelar aspectos importantes da atividade científica do país. Sustenta tal hipótese o fato de serem esses trabalhos acadêmicos produzidos e avaliados em instâncias altamente organizadas e controladas - os programas de pós-graduação.

Para testar semelhante hipótese, foram analisadas as dissertações e teses de duas áreas distintas do conhecimento: Pesquisa Energética e Nuclear e Ciência da Informação. Os procedimentos metodológicos relativos à constituição do *corpus* e seu tratamento estão apresentados no item a seguir.

2. Metodologia

A produção de indicadores da ciência, por métodos bibliométricos, requer um conjunto de dados padronizados, sistematizados e consistentes, em princípio encontráveis nas bases de dados bibliográficos. Essas fontes favorecem as abordagens bibliométricas em razão da forma em que os dados se encontram estruturados. No entanto, na prática, as inconsistências encontradas impedem sua utilização direta, sendo necessário realizar um trabalho prévio de consistência e reestruturação das bases. Para isso, recorre-se a procedimentos e *softwares* bibliométricos criados especificamente para promover a consistência requerida.

Neste trabalho, o *corpus* foi constituído de três tipos de dados: a) as dissertações e teses defendidas, nos últimos 30 anos (1978-2005), no Programa de Pós-graduação do Instituto de Pesquisas Energéticas e Nucleares; b) os projetos de pesquisa em andamento, apresentados a esse mesmo Programa (2001 a 2006) e c) as dissertações e teses de Ciência da Informação produzidas nos programas nacionais de pós-graduação (1977 a 2001).

Tendo em vista a natureza deste estudo, focada na descrição temática das dissertações e teses, procedeu-se, inicialmente, à análise dos descritores atribuídos a cada item das bases de dados consideradas. Os descritores foram reorganizados para se obter um conjunto normalizado quanto à forma e ao conteúdo.

Para a obtenção de conjuntos homogêneos, os descritores foram integrados a categorias temáticas mais abrangentes. Contornou-se, assim, a dispersão provocada pelos descritores originalmente atribuídos. Estudos anteriores demonstraram que as operações de reorganização temática são necessárias para garantir a confiabilidade dos resultados. Dito de outro modo, a aplicação de critérios apenas estatísticos à análise temática promove a eliminação de parte significativa dos dados, o que torna o conjunto final pouco representativo.

Para garantir o rigor das operações recorreu-se aos princípios propostos pela área de Organização do Conhecimento (Kobashi; Santos, 2006a e 2006b), com a utilização dos *softwares* *Dataview*¹ e *Infotrans*². Em síntese, os dados do *corpus* foram submetidos a processos de consistência e posterior armazenamento em bases de dados *ad hoc*.

3. Análise das dissertações e teses do Instituto de Pesquisas Energéticas e Nucleares - IPEN-CNEN/SP

O Instituto de Pesquisas Energéticas e Nucleares (IPEN) é uma autarquia pública gerida administrativamente pela Comissão Nacional de Energia Nuclear, instituição federal subordinada ao

¹ Software francês, produzido pelo Centre de la Recherche Retrospective de Marseille (França), utilizado para gerar listas de frequência e matrizes de co-ocorrência sob a forma de tabelas e gráficos (Kobashi; Santos, 2006a).

² Software alemão, da empresa LuK (Information und Kommunikation), utilizado para transformar registros bibliográficos em registros bibliométricos, padronização automática de dados, criação de novos campos e reorganização geral da Base de dados (Idem).

Ministério de Ciência e Tecnologia. A instituição mantém, há 30 anos, um curso de pós-graduação em *Tecnologia Nuclear*, avaliado com nível 6 pela CAPES.

No âmbito do IPEN, a unidade de informação, denominada Biblioteca *Terezine Arantes Ferraz*, é responsável pelos serviços de apoio à pesquisa, gerenciamento do acervo institucional e criação e manutenção de bases de dados de produção científica. A memória científica da instituição é preservada e gerenciada por meio de uma base de dados referencial denominada, Produção Técnico-Científica – *PTC*, estruturada em padrões internacionais, semelhante à base de dados do *INIS* – *International Nuclear Information System*, da *International Atomic Energy Agency* – *IAEA* (Viena, Áustria).

A atribuição de metadados, bem como a indexação dos documentos desse repositório, seguem padrões e políticas estabelecidos pelo *INIS* para todos os países membros da *Internacional Atomic Energy Agency- IAEA*. Para utilizar, de forma criteriosa, os referidos padrões e políticas, os profissionais do IPEN, encarregados das tarefas de tratamento da informação, receberam capacitação específica que inclui conhecimentos sobre os fundamentos da área nuclear e sobre a aplicação da metodologia *INIS*. Dessa forma, os dados, tratados tecnicamente por meio de um padrão consolidado, apresentam um grau bastante satisfatório de qualidade, adequados para a aplicação de técnicas métricas, demandando apenas pequenos ajustes para obter a consistência final requerida.

Da base de dados *PTC* foram selecionadas as dissertações e teses produzidas de 1976 a 2005 (num total de 1190 itens), que passaram por um processo de verificação de consistência para posterior armazenamento em uma nova base de dados, conforme explicitado no item 2 acima. O *corpus* foi enriquecido com dois campos suplementares: Áreas de concentração do Programa de Pós-graduação e Categoria de assuntos, para o qual foi utilizado como parâmetro de categorização o *Subject Categories and Scope Description* do *INIS*.

A FIG. 1, abaixo, mostra a distribuição temática das dissertações e teses produzidas pelo Programa de Pós –graduação do IPEN. A denominação das áreas foi mantida em inglês, respeitando-se a terminologia da *Subject categories and scope descriptions* do *INIS*.

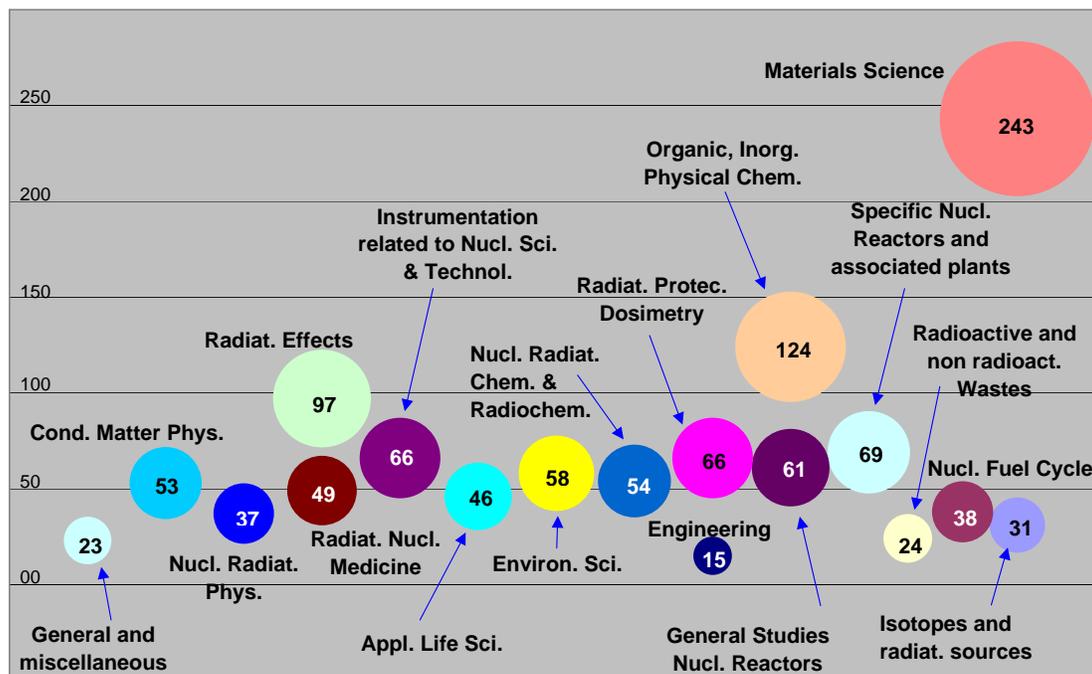


FIGURA 1 – Distribuição temática das dissertações e teses do Programa de Pós-graduação do IPEN (1976-2005)

Esta figura mostra a prevalência de três grandes áreas ao longo do período estudado: a) *Materials Science*, b) *Organic, Inorganic and Physical Chemistry* e c) *Radiation Effects*. Na visão do especialista, que interpretou os dados da FIG. 1, os resultados são coerentes com os aspectos políticos e sociais do país, acontecidos no decorrer do período, em particular com a política científica nacional e com as decisões tomadas pelo Programa de pós-graduação. Além desta visualização, foram obtidos também outros resultados, como, por exemplo, a evolução temática por quinquênios, o surgimento de novas áreas de pesquisa no decorrer do período, a dinâmica das áreas, que ora ocupam as primeiras posições no *ranking* de produção de dissertações, ora são ultrapassadas por outras.

Para os fins da elaboração da política científica institucional, além de conhecer o percurso da pesquisa (portanto, o passado), é útil identificar as tendências do futuro. Para conhecer as temáticas que se delineiam para os próximos anos, foram analisados os projetos de dissertações e teses apresentados ao Programa no período 2001-2006, totalizando 405 projetos. Para essa análise, foi construída uma nova base de dados de projetos em andamento, cuja estruturação e categorização é semelhante à da base de dissertações e teses. Assim, foi possível visualizar não só os temas das pesquisas em andamento, como também a sua distribuição no próximo quinquênio. As FIG. 2 e 3, a seguir, apresentam, de forma resumida, um prognóstico dos temas das dissertações e teses que serão defendidas até 2010.

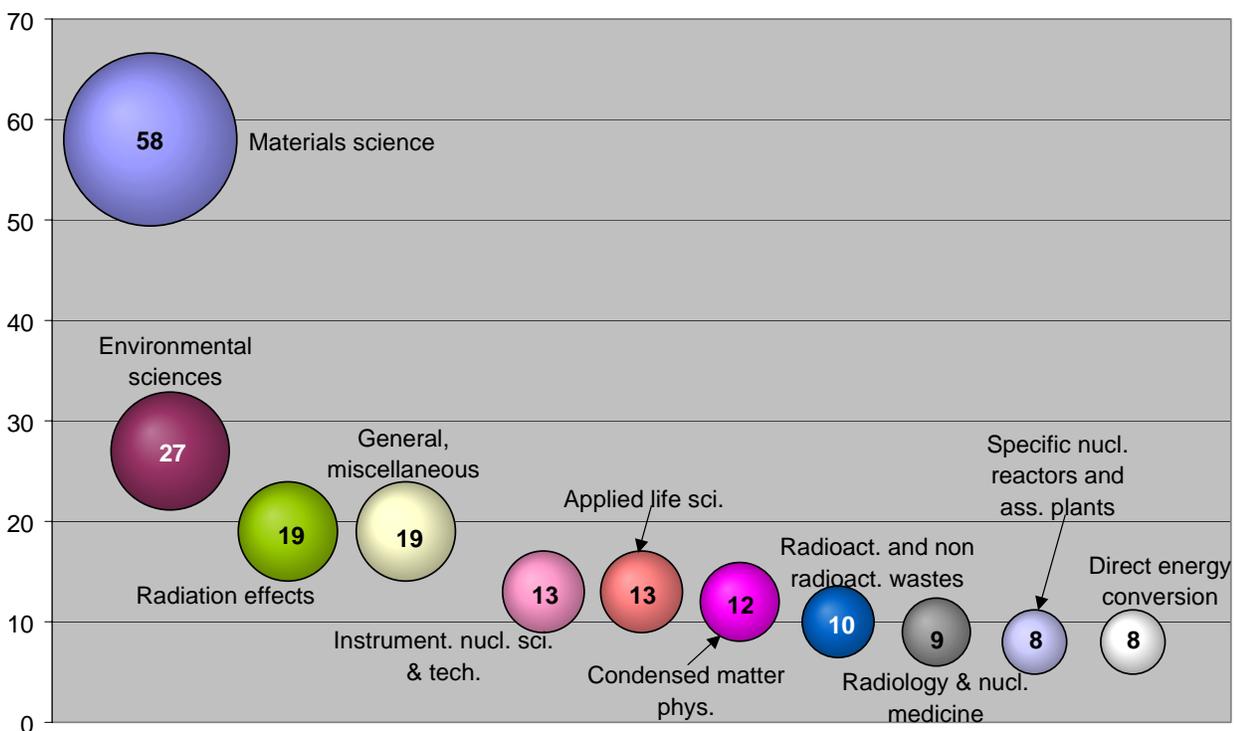


FIGURA 2 – Temas dos projetos de dissertações de Mestrado que serão defendidos até 2010

A FIG. 2, relativa às dissertações de mestrado em andamento, mostra que estão ocorrendo algumas mudanças se comparados aos dados relativos às dissertações e teses já defendidas nos 30 anos anteriores, apresentadas na FIG. 1. Nota-se o significativo crescimento de projetos voltados à

área ambiental (*Environmental Sciences*), resultado que é fruto de decisão adotada pela direção do IPEN, há uma década atrás, dentro de seu planejamento estratégico, que introduziu no Programa de Pós-graduação às preocupações globais com os problemas ambientais; evidencia-se também o crescimento da categoria *General e miscellaneous*, na qual estão integrados os estudos da área de Gestão do conhecimento, abordagens importantes para o gerenciamento das pesquisas.

Com relação aos projetos de doutorado, FIG.3, registra-se o aparecimento de estudos voltados para Energias Alternativas, uma linha de pesquisa considerada importante na missão do Instituto. A constante predominância do tema Ciência dos Materiais (*Materials Science*), por sua vez, está relacionada à natureza interdisciplinar desse campo de estudos.

Com efeito, o estudo dos materiais está presente em todas as linhas de pesquisa da instituição e contribui também para esta predominância a densidade da massa crítica de pesquisadores dessa área ora em atividade no IPEN. Comportamento semelhante apresenta o tema Efeitos das radiações (*Radiation effects*), área também interdisciplinar, em que são desenvolvidos estudos sobre o uso da radiação para aplicações médicas, esterilização de materiais, preservação de alimentos, proteção de plantações, entre outros

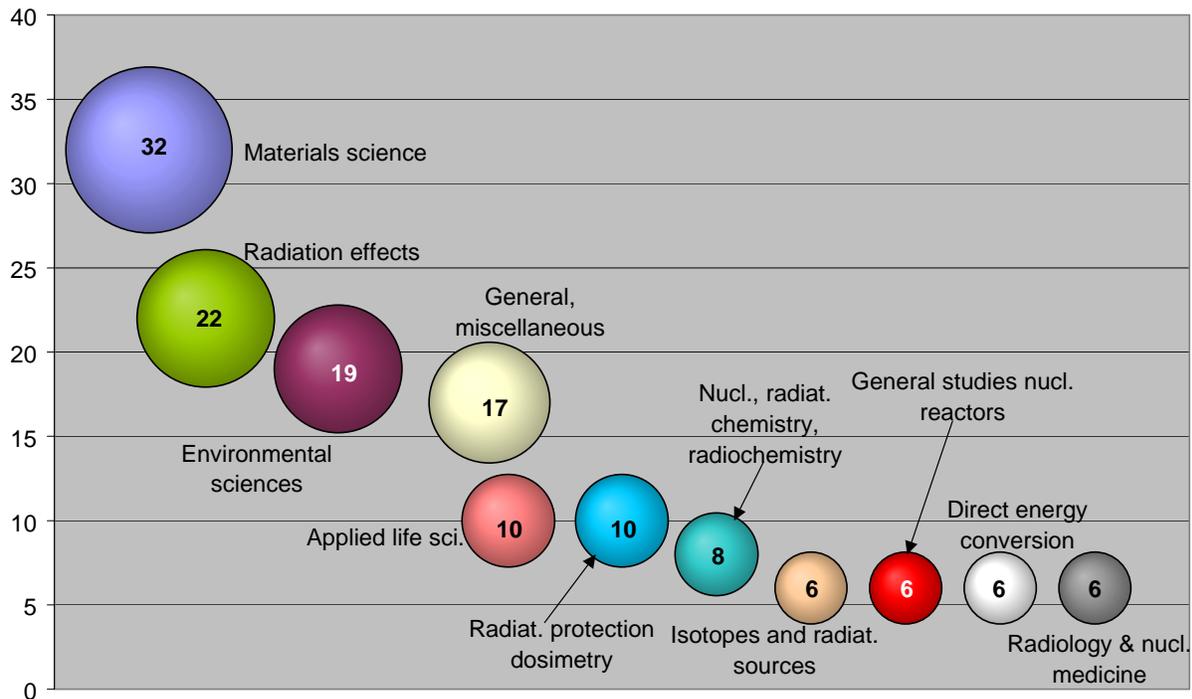


FIGURA 3- Temas de projetos de doutorado que serão defendidos até 2010

4. Análise das dissertações e teses de Ciência da Informação

A área da Ciência da Informação, no Brasil, conta atualmente com nove programas de pós-graduação credenciados pela CAPES. Foi considerado para estudo o período 1978 a 2001. Dois programas foram criados e credenciados recentemente, não existindo, portanto, dados referentes ao período especificado. Deve-se assinalar que os dados de Ciência da Informação, posteriores a 2001, por serem assistemáticos e parciais, não possibilitaram estender o período de cobertura do estudo. Observou-se, também, o alto grau de desorganização, dispersão e ausência de padronização das bases de dados dessa área. Para garantir a confiabilidade do estudo, foi necessário realizar um exaustivo trabalho complementar de coleta e padronização dos dados³.

A análise aqui realizada refere-se às teses e dissertações defendidas nos programas de Ciência da Informação das seguintes instituições: Universidades de Brasília, Universidade Federal de Minas Gerais, IBICT/Universidade Federal Fluminense, Universidade Federal da Paraíba, Universidade Estadual de São Paulo, Universidade de São Paulo e Pontifícia Universidade Católica de Campinas. Esses Programas têm nota que varia de 3 a 5, segundo os critérios da CAPES, sendo, portanto, área ainda não consolidada.

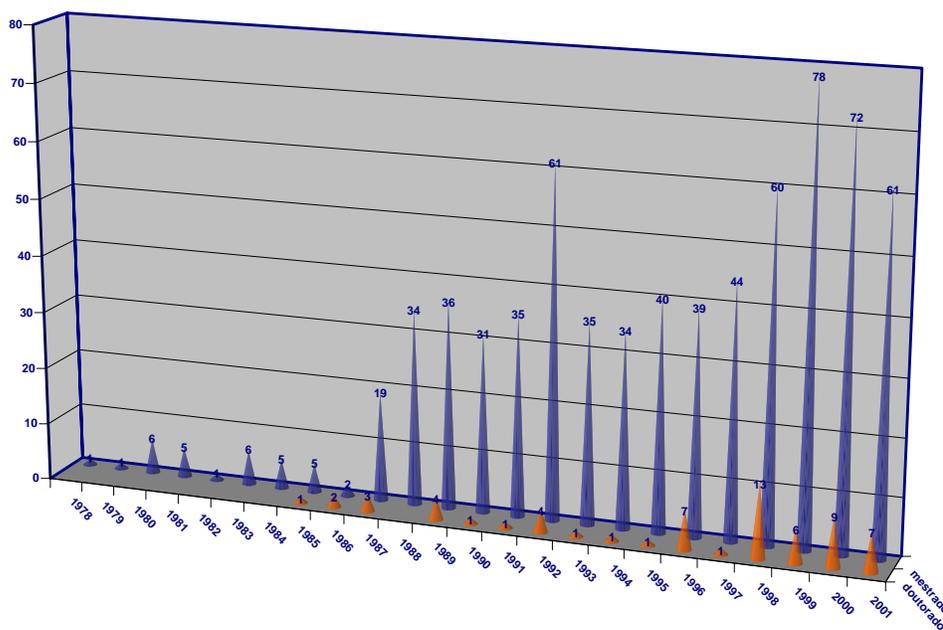


FIGURA 4 - Total de dissertações e teses de Ciência da Informação, por ano (1978-2001).

A FIG. 4 apresenta o número total de teses e dissertações defendidas nos Programas de Pós-graduação em Ciência da Informação do país, no período de 1978-2001. De um total de 833 trabalhos, 749 são dissertações de mestrado e 84 são teses de doutorado. Chama a atenção a quantidade significativamente menor de teses de doutorado (em torno de 11% do total), em relação às dissertações de mestrado.

³ A coleta e análise de dados referente ao período 2002-2006 está em processo. Pretende-se apresentar os resultados ainda no presente ano.

Esse resultado é preocupante, na medida em que a consolidação de um campo está fortemente relacionada ao número de doutores. Confirma-se, desse modo, o que a CAPES já havia detectado na avaliação trienal de 2003: a área da Ciência da Informação não tem tido crescimento compatível com o que se verifica nas outras áreas das Ciências Sociais Aplicadas.

Um outro aspecto estudado refere-se à institucionalização cognitiva da área, que pode ser identificada pelos descritores atribuídos aos trabalhos. Na base de dados original, para 833 documentos, foram atribuídos um total de 1929 descritores. Por ser considerável a variação do número de descritores atribuídos a cada documento, foram selecionados para teste os 33 descritores utilizados com frequência igual ou maior a 10. Esses 33 descritores (2% do total) foram atribuídos a 78% do total dos documentos. A FIG. 5, abaixo, apresenta a distribuição de descritores das dissertações e teses.

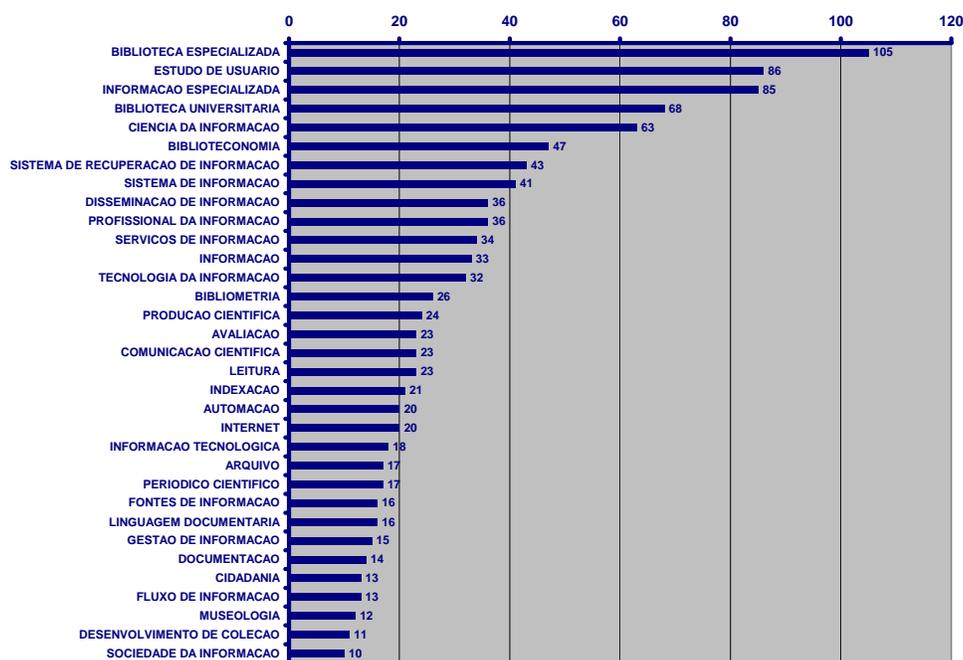


Figura 5 - Descritores atribuídos pelos autores das dissertações e teses.

Os resultados acima foram obtidos pela contagem dos descritores atribuídos pelos autores das dissertações e teses. Observa-se que termos como “Produção científica” e “Comunicação científica” podem, em certos contextos, representar um mesmo processo. “Informação especializada” e “Informação tecnológica” podem ser também consideradas expressões sinônimas de um mesmo conceito. A ausência de padronização gramatical dos descritores, aliada à não utilização de um vocabulário controlado de referência, para indexação, como pode ser observado na figura acima, e pelo resultado da análise da matriz de similaridade, realizada a partir do estudo de co-ocorrências desses mesmos descritores, não fornece, portanto, informações suficientes para a realização de inferências confiáveis.

Para o desenvolvimento do estudo, estabeleceu-se uma metodologia de padronização dos elementos bibliográficos de interesse e produção de uma base de dados *ad hoc*. Observou-se que o

uso simultâneo de 2 descritores (co-ocorrência entre esses mesmos 33 descritores) ocorre somente em 357 documentos, ou seja, em menos de 43% do total de documentos⁴.

Nesse sentido, foram criados novos campos, como segue: a) **campo de descritores**: para registrar os descritores atribuídos pelo autor, porém padronizados quanto à forma; b) **campo de macro-descritores**: neste caso, cada documento recebeu dois descritores. O descritor de nível mais genérico refere-se aos nomes dos Grupos de Trabalho (GTs) da ANCIB (Associação Nacional de Pesquisa e Pós-graduação em Ciência da Informação⁵, apresentados na FIG. 7. A cada categoria foram subsumados descritores específicos retirados da ementa de cada grupo (FIG. 6).

A adoção das denominações dos GTs da Ancib foi considerada pertinente por serem eles instâncias organizadoras das discussões promovidas pela associação, como também de distribuição temática dos trabalhos nos encontros anuais da entidade. São, portanto, categorias temáticas reconhecidas e legitimadas pelos pesquisadores da Ciência da Informação do país.

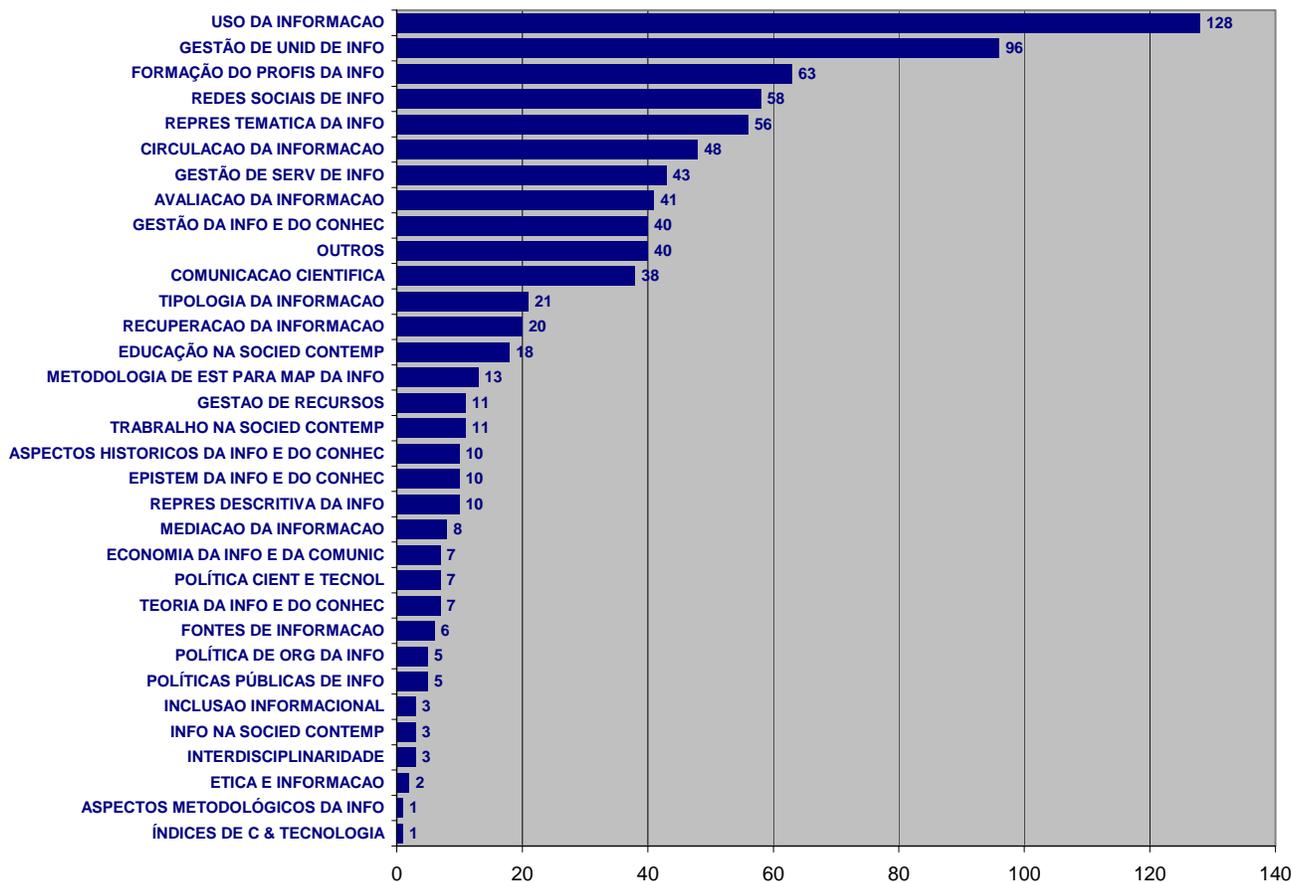


FIGURA 6 – Redistribuição das dissertações e teses segundo termos extraídos das ementas dos Grupos de Trabalho da ANCIB.

⁴ Dessa forma, um estudo posterior, por meio de construção de matriz de similaridade, *input* para análise de *clusters* (Van Raan, 2004), que permite obter inferências sobre a institucionalização cognitiva da área, será realizado com esses 357 documentos.

⁵ Mediação, circulação e uso da informação; 2) Gestão de unidades de informação; 3) Organização do conhecimento e representação da informação; 4) Informação e trabalho; 5) Informação para diagnóstico, mapeamento e avaliação; 6) Estudos históricos e epistemológicos da informação; 7) Política, ética e economia da informação.

A FIG. 6 mostra a distribuição das dissertações e teses obtida por reindexação, utilizando-se como referência os termos constantes da descrição de cada ementa dos GTs. Obtém-se, com este procedimento, uma visão mais acurada do perfil da área porque a utilização de um vocabulário controlado de referência evita as inconsistências e dispersão provocada pela atribuição de descritores pelo próprio autor da dissertação ou tese.

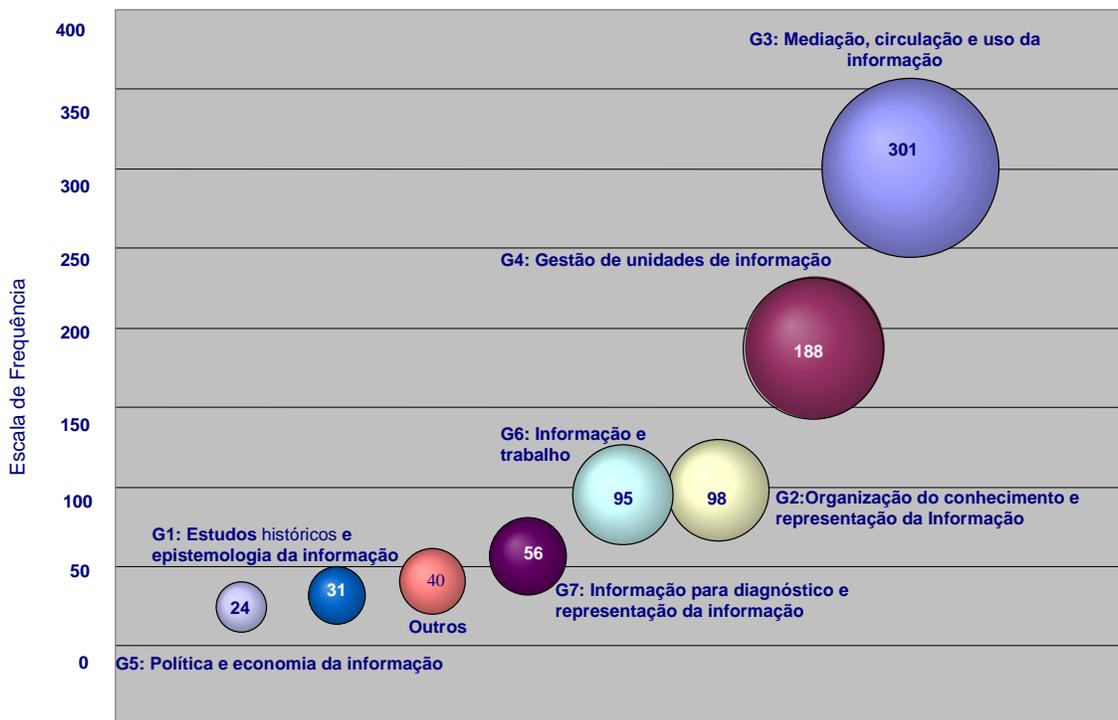


Figura 7 - Distribuição de dissertações e teses de Ciência da Informação (1978-2001), segundo categorias dos Grupos de Trabalho da ANCIB.

A FIG. 7 mostra um grau maior de generalização na distribuição das 834 teses do que a da FIG.6. A categoria que concentra o maior número de trabalhos é Mediação, Circulação e Uso da informação, com 301 itens; vêm, em seguida, as categorias Gestão de Unidades de Informação (188 itens), Organização do Conhecimento (98 itens) e Informação e Trabalho (95 itens). As categorias Informação para Diagnóstico, Estudos Históricos e Epistemológicos e Política e Ética da Informação são as que apresentam menor quantidade de itens. Na categoria Outros foram incluídas as dissertações e teses que não se encaixam em nenhuma dessas 7 categorias.

A distribuição acima permite fazer diferentes questionamentos: O GT3, ao contemplar praticamente 1/3 do total de trabalhos da área, é de fato o núcleo da área ou, ao contrário, aí se concentram trabalhos de natureza genérica, ou são trabalhos interdisciplinares por excelência? A presença tímida de trabalhos de natureza epistemológica e histórica (GT1) e sobre as questões políticas, econômicas e éticas revelam a pouca preocupação da área com esses temas? Ou, ao contrário, essas questões estão explicitamente incorporadas às teses e dissertações, devendo-se adotar outras técnicas de análise para revelá-las? Semelhantes indagações que requerem aprofundamento, poderão, certamente, ser objetos de novas pesquisas da área.

5. Considerações finais

Os resultados apresentados neste estudo podem subsidiar um processo maior e mais abrangente de descrição e avaliação da institucionalização dos campos científicos para o planejamento de políticas científicas. A atividade científica implica a inversão de recursos públicos, devendo, portanto, ser objeto de avaliações contínuas por métodos confiáveis. Nessa medida, a avaliação, para ser eficaz, deve associar aspectos quantitativos e qualitativos, fato que requer teoria e instrumentos complexos. Procurou-se, neste trabalho, demonstrar que o uso de técnicas bibliométricas e cienciométricas são úteis porém, análises e interpretações acríticas, podem gerar resultados pouco significativos.

Chamam a atenção, no estudo, as diferenças nos graus de institucionalização entre as áreas da Ciência da informação e Energia Nuclear. De fato, a Ciência da Informação, como um todo, produziu 833 dissertações e teses no período de 1978 a 2001, enquanto um único programa da área de Energia Nuclear (IPEN) produziu, em 30 anos, 1190 dissertações e teses. Essa configuração decorre do grau de amadurecimento das áreas, fenômeno que pode ser atribuído ao grau de incentivo efetivamente recebido pelas áreas, nos planos de desenvolvimento da Ciência e da Tecnologia do país. A área da Ciência da informação não é área de conhecimento considerada prioritária nessas políticas, por não ser vista como campo de pesquisa estratégica.

Por outro lado, os inputs financeiros destinados pelas agências de financiamento à pesquisa e a configuração da massa crítica de cada área podem ter impactos perversos na infra-estrutura de pesquisa. As bases de dados de produção científica são parte dessa infra-estrutura. Sem bases de dados exaustivas, elaboradas com rigor, como se procurou demonstrar, ficam comprometidos os estudos de produção de parte significativa de indicadores da ciência.

As bases de dados bibliográficos elaborados, via de regra, por unidades de informação, constituem matéria-prima básica para o desenvolvimento de estudos de Produção Científica. Construídas com o principal objetivo de facilitar a busca e recuperação de informação, essas bases vêm se configurando como dispositivos multifuncionais que podem ser utilizados para estudos de atividade científica, gestão de acervos, análise de produtividade de pesquisadores, redes de colaboração, entre outras funções. Em todos esses usos, é desejável que os repositórios sejam consistentes, portanto confiáveis.

Observou-se que os repositórios de Ciência da Informação não apresentam a mesma qualidade e consistência dos do IPEN. Alguns fatores podem explicar as diferenças de qualidade. No caso da Ciência da Informação, os dados provêm fundamentalmente da Base de Dissertações e Teses da CAPES. Como se sabe, os dados dessa base são fornecidos pelas Programas de Pós-graduação do país, não havendo, nem na fonte geradora, nem na instituição receptora, um trabalho de consistência dos dados. No caso do IPEN, a consistência resulta da política de tratamento de informação adotada pela unidade de informação da instituição, que é responsável pela indexação da produção científica local, segundo critérios e linguagem padronizados internacionalmente.

O estudo ora realizado indicou que a produção de indicadores necessita de procedimentos prévios de consistência, de modo a adequá-los às análises bibliométricas. Indicou, por outro lado, a necessidade de investimento na formação de recursos humanos aptos a produzir bases de dados para as diferentes finalidades acima especificadas. Dentre os conhecimentos requeridos podem ser citados ao menos: *a)* domínio de técnicas de construção de bases de dados com estrutura de metadados aptas a responder a diferentes funcionalidades; *b)* sólidos conhecimentos relativos à organização de conhecimento (tratamento de conteúdo e instrumentos de representação do conhecimento); *c)* avaliação de qualidade de bases de dados (técnicas quantitativas e qualitativas).

O trabalho multidisciplinar é também fundamental para o sucesso dos estudos de produção científica. A participação de diversos “atores”, com pluralidade de conhecimentos, agrega valor ao estudo. Neste trabalho, associaram-se especialistas de cada campo do conhecimento analisado, especialistas em organização e estruturação de bases de dados, estudiosos de métodos bibliométricos e de organização do conhecimento. No caso específico da Energia Nuclear, os resultados foram analisados e interpretados pelo gestor de P&D do IPEN, à luz do contexto social,

político e econômico e de seus impactos nas políticas de pesquisa do Programa de pós-graduação da instituição. Os dados de Ciência da Informação foram também objeto de análise de pesquisadores dessa área.

Alguns fatores, portanto, devem ser apontados como pré-requisitos dos estudos de Produção Científica: *a)* a qualidade dos dados a serem analisados, tais como: completeza, consistência, disponibilidade, abrangência; *b)* teorias e métodos consistentes que sustentem o processo de análise dos dados; *c)* disponibilidade de softwares bibliométricos; *d)* trabalho em equipe. Não seria demais observar que esses pré-requisitos são largamente enunciados, porém pouco adotados na prática.

Por último, foi possível comprovar a potencialidade das técnicas bibliométricas tanto para elaborar retratos de produção científica já concluída, quanto para elaborar prognósticos, como no caso dos estudos dos projetos em andamento no IPEN. Retratos e prognósticos são, certamente, informação estratégica útil para subsidiar a tomada de decisão de gestores de política científica. Porém, a construção de retratos e prognósticos é refratária à improvisação.

6. Bibliografia

IGAMI, Mery P. Z. ; NOZAKI, Priscila; KOBASHI, Nair Y. A gestão de qualidade em bases de dados da produção científica In: Congresso brasileiro de biblioteconomia, documentação e ciência da informação, 12, 8-11 julho, 2007, Brasília, DF. (Trabalho aceito para apresentação).

INTERNATIONAL NUCLEAR INFORMATION SYSTEM – INIS. **Joint thesaurus**. Vienna: International Atomic Energy Agency, 2004. (ETDE/INIS Joint reference series, nº 1, rev. 1)

INTERNATIONAL NUCLEAR INFORMATION SYSTEM – INIS. **Subject categories and scope descriptions**. Vienna: International Atomic Energy Agency, 2002. (ETDE/INIS Joint reference series nº 2).

KOBASHI, Nair Y.; SANTOS, Raimundo N. M. dos. Arqueologia do trabalho imaterial: uma aplicação bibliométrica à análise de dissertações e teses. In: Encontro Nacional de Pesquisa em Ciência da Informação – 7.,2006, Marília. **Anais....** Marília: FFC/UNESP, 2006b. 1 CD ROM. (VII ENANCIB)

KOBASHI, Nair Y.; SANTOS, Raimundo N. M. dos. Institucionalização da pesquisa científica no Brasil: cartografia temática e de meios sociais pôr meio de técnicas bibliométricas. **Transinformação**, Campinas, v. 18, n. 1, p. 27-36, jan./abr., 2006a.

MACIAS- CHAPULA, Cesar A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ciência da Informação**, vol.27, n.2 n.p., 1998. Disponível em: < <http://www.scielo.bireme.br> > acesso em 02 maio 2007.

SANTOS, Raimundo N.M. dos; KOBASHI, Nair Y.; IGAMI, Mery P. Z. BRESSIANI, José C. Brazil's insertion into nuclear energy research: bibliometric analysis of dissertations developed at the Instituto de Pesquisas Nucleares- São Paulo. In: GUERREIRO-BOTE, Vicente P. (Ed.). Current research in information sciences and technologies. Multidisciplinary approaches to global information systems . International conference on multidisciplinary information sciences and technologies, 1st, october 25-28th, 2006, Merida Spain. **Proceedings...** Badajoz, España: Open Institute of Knowledge, 2006. V.1, p.553-556.

SPINAK, Ernesto Scientometric indicators. *Ciência da Informação*, vol.27, no.2, n.p., 1998. Disponível em < [http:// www.scielo.bireme.br](http://www.scielo.bireme.br)> acesso em : 02 maio, 2007.

TAGUE-SUTCLIFFE, J. An introduction to informetrics. *Information Processing & Management*. v. 28, n. 1, p. 1-3, 1992. Disponível em : < <http://www.periodicos.capes.gov.br> > acesso em: 02 maio, 2007.

Van RAAN, Anthony .F.J. The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments. Pre-print, 2003. Disponível em <http://www.cwts.nl/TvR/documents/AvR-TFA2003.pdf> < acesso em: maio 2007.

Agradecimentos: Nossos mais profundos agradecimentos à bolsista de Iniciação Científica Priscila Nozaki, aluna do Curso de Biblioteconomia da ECA-USP, pelo trabalho realizado na organização e formatação dos dados. Sua contribuição tem sido fundamental para a realização destes estudos. Agradecemos ao CNPq pelos auxílios à pesquisa recebidos e à Pró-Reitoria de Pesquisa da USP pela concessão da Bolsa Especial a Alunos de Graduação.