

Área: Gestão

Sub-área temática: Qualidade e produtividade de serviços em unidades de informação

A GESTÃO DE QUALIDADE EM BASES DE DADOS DA PRODUÇÃO CIENTÍFICA

MERY P. Z. IGAMI*

PRISCILA NOZAKI**

NAIR Y. KOBASHI***

Resumo

A qualidade da recuperação da informação, em bases de dados, está diretamente relacionada à qualidade dos registros. As bases de dados vêm sendo, também, crescentemente utilizadas para a produção de indicadores bibliométricos, em vista da forma de estruturação dos dados. Antecede a análise bibliométrica a verificação de consistência dos registros bibliográficos. Nessa medida, propõe-se a utilização de softwares bibliométricos para a gestão de qualidade das bases de dados bibliográficos, tanto para promover melhor recuperação de informação, quanto para a produção de indicadores confiáveis.

Palavras-chave: Gestão da informação. Gestão de qualidade de bases de dados. Bibliometria. Cientometria.

* Bibliotecária chefe da Divisão de Informação e Documentação Científicas, IPEN-CNEN/SP. Av. Lineu Prestes 2242, Butantã, CEP: 05508-000, São Paulo, SP – Brasil. E-mail: mery@ipen.br

** Estudante de graduação do curso de Biblioteconomia e Documentação, ECA-Universidade de São Paulo (aluna de iniciação científica). Av. Prof. Lúcio Martins Rodrigues, 443, Butantã, CEP: 05508-900, São Paulo, SP – Brasil. E-mail: prinozaki@yahoo.com.br

*** Professora livre-docente, ECA-Universidade de São Paulo - Av. Prof. Lúcio Martins Rodrigues, 443, Butantã, CEP: 05508-900, São Paulo, SP – Brasil. E-mail: nykobas@usp.br

1. Introdução

A qualidade das bases de dados é influenciada por vários fatores: políticas de tratamento da informação, *softwares* de armazenamento e de busca, interfaces de suporte ao usuário, *hardware* e telecomunicações (PEREIRA et al, 1999).

Um dos aspectos qualitativos fundamentais das bases de dados refere-se ao seu conteúdo. Uma base de dados com erros pode ocasionar prejuízos, não só intelectuais e financeiros, bem como comprometer a credibilidade do serviço de informação, como um todo, junto aos usuários finais. Os mecanismos de coleta e registro são dimensões importantes da qualidade do conteúdo das bases. Atividades como planejamento de estratégia de crescimento da base, mecanismos de limpeza dos registros e reorganização de conteúdo são fatores críticos de sucesso e necessários para a manutenção das bases.

O tema gestão de qualidade em bases de dados atraiu a atenção da indústria da informação apenas na segunda metade da década de 90. Os trabalhos apresentados estavam relacionados com a detecção e correção de erros de ortografia e a aplicação de processos de autochecagem de dados legíveis por máquina. Os estudos tinham como alvo principal as bases de dados bibliográficos, em decorrência do desenvolvimento de grandes coleções gerenciadas por meios eletrônicos, cujo início remonta aos anos 60 (PEREIRA et al, 1999). Embora fundamental para a garantia de qualidade das bases, observa-se que poucos foram os métodos de gestão de qualidade efetivamente propostos e adotados.

A utilização de bases de dados bibliográficos em estudos bibliométricos é uma prática consagrada. A consistência dos dados de partida é, sob essa ótica, fundamental para a obtenção de resultados confiáveis. De fato, os estudos bibliométricos encontram, nesses repositórios, dados estruturados de forma adequada para serem classificados e quantificados (ROSTAINING, 1996). No entanto, muitas vezes, sua utilização para fins bibliométricos requer um trabalho anterior de consistência dos registros. Alguns procedimentos e *softwares* bibliométricos foram criados exatamente para obter a consistência requerida.

Este trabalho procura chamar a atenção para a importância da utilização dos métodos e *softwares* bibliométricos na gestão de qualidade de bases de dados bibliográficos. Esta perspectiva ficou evidenciada quando da utilização da Base de dados de dissertações e teses do IPEN/CNEN/SP para a análise das tendências de pesquisa do Programa de Pós-graduação da instituição, para subsidiar a política de pesquisa do programa. A utilização de mecanismos de verificação e checagem automática de conteúdos, bem como de recursos métricos de avaliação, que garantam a melhoria contínua das bases, mostrou ser parte imprescindível da gestão de qualidade de bases de dados.

Para melhor compreensão da experiência realizada, será inicialmente caracterizada a instituição em que foi efetuado o estudo e, em seguida, serão apresentados a metodologia, os instrumentos utilizados e parte dos resultados obtidos.

2. Caracterização da instituição

Fundado em 1956, o Instituto de Pesquisas Energéticas e Nucleares em São Paulo (IPEN) é uma autarquia federal cuja missão é: *Contribuir com a melhoria da qualidade de vida da população brasileira, produzindo conhecimentos científicos, desenvolvendo de tecnologias, gerando produtos e serviços e formando recursos humanos nas áreas nuclear e correlatas* (Plano Diretor – IPEN 2005).

As atividades do IPEN distribuem-se em: pesquisa e desenvolvimento, ensino e produção. Assim sendo, um dos principais produtos intelectuais da sua comunidade é a Produção Científica, constituída de trabalhos apresentados em eventos científicos, artigos publicados em periódicos tanto em âmbito nacional como internacional, livros, dissertações, teses e relatórios científicos.

Com relação às atividades de ensino, o IPEN mantém, há 30 anos, um curso de pós-graduação em *Tecnologia Nuclear*, avaliado com nível 6 pela CAPES. A cada ano são produzidas em média 120 novas dissertações e teses.

No âmbito do IPEN, a unidade de informação, denominada *Biblioteca Terezine Arantes Ferraz*, foi criada para fornecer suporte bibliográfico e serviços

de informação especializados à comunidade do instituto. O acervo da biblioteca é especializado na área nuclear e ciências relacionadas. Para gerenciar este acervo, a unidade mantém bases de dados e catálogos acessíveis *on-line*, nos quais se apoia para realizar seus principais serviços de informação. Para assegurar a alta qualidade dos serviços de informação prestados, todos os profissionais que nela atuam são permanentemente capacitados, tanto em relação ao conhecimento generalista da área nuclear, quanto nos conhecimentos teóricos e práticos necessários ao tratamento da informação e manutenção das bases de dados.

2.1. Base de dados da Produção Técnico - Científica - PTC

Uma das missões da Biblioteca da instituição é preservar a memória da produção científica da instituição. Para tanto, mantém uma base de dados referencial estruturada em padrões internacionais, semelhante à base de dados do *INIS – International Nuclear Information System*, produzida pela *International Atomic Energy Agency – IAEA*, localizada em Viena, na Áustria. A base de dados do *IAEA* contém, atualmente, aproximadamente 2,8 milhões de itens bibliográficos da produção científica mundial sobre energia nuclear e ciências relacionadas. Essa base está disponível para consulta *on-line*, gratuita, em todos os países membros da *IAEA*.

Na base de dados de produção científica do IPEN, denominada Produção Técnico-Científica – *PTC*, são registrados os trabalhos apresentados em eventos nacionais e internacionais, artigos publicados em periódicos, as dissertações e teses produzidas pelo curso de pós-graduação, bem como outros documentos produzidos pela comunidade técnico-científica do IPEN.

Desde sua criação, esta base de dados atende a diversos objetivos, quais sejam: *a)* preservar a produção técnico-científica da instituição; *b)* conferir visibilidade à produção técnico-científica. Os documentos que integram a base também alimentam a base internacional do *INIS*; *c)* fornecer dados estatísticos para a direção do instituto; *d)* servir como fonte de consulta primária para alunos de pós-graduação iniciando os seus trabalhos de pesquisa e reconhecimento da área.

Atualmente, a base *PTC* contém aproximadamente 11.600 registros. Encontra-se representado nessa base de dados o conhecimento produzido pela comunidade científica do IPEN, desde a criação da instituição. Trata-se, portanto, de um rico repositório que contém matéria-prima imprescindível para dar suporte à pesquisa científica. Essa mesma base de dados, se submetida a diferentes tipos de análises, pode gerar indicadores importantes para a tomada de decisão e para a formulação de política de pesquisa científica institucional.

A base *PTC* vem sendo utilizada tradicionalmente para gerar dados estatísticos de produção científica. Em 2005 foi realizado um estudo objetivando conhecer a produtividade dos pesquisadores do IPEN com relação aos artigos publicados em revistas internacionais (IGAMI, 2005). No entanto, a base não havia sido utilizada, ainda, de forma extensiva, para a produção de indicadores da ciência, tais como o perfil temático da pesquisa e a colaboração entre pesquisadores.

Em 2006, deu-se início a um projeto de exploração sistemática da base *PTC* por métodos bibliométricos. É essa experiência que será aqui relatada. Por ser estudo exploratório, foi selecionada apenas parte do conteúdo da base, especificamente os dados sobre as dissertações e teses produzidas no Programa de Pós-graduação do IPEN.

3. Metodologia

A produção de indicadores, por métodos bibliométricos, requer conjuntos de dados padronizados, sistemáticos e consistentes. Com efeito, as bases de dados bibliográficos são largamente utilizadas em estudos bibliométricos exatamente porque apresentam, em princípio, as qualidades requeridas.

Em décadas anteriores, os estudos bibliométricos estavam fortemente concentrados na gestão de acervos, mormente os estudos utilizassem as conhecidas leis de Bradford, Lotka e Zipf. Foi na última década que, favorecida pelo expressivo desenvolvimento das tecnologias da informação, a bibliometria transformou-se em um valioso instrumento não só para gestão de acervo, mas principalmente para a gestão de conteúdo. Nesse mesmo período, surgem outras

aplicações da bibliometria, cujas denominações, via de regra, estão associadas à mídia utilizada: *webometria*, *informetria*, por exemplo, sendo, porém, pouco precisa a conceituação desses termos.

De acordo com a definição de Tague-Sutcliffe (1992), a bibliometria refere-se ao estudo dos aspectos quantitativos da produção, disseminação e uso da informação explícita. A bibliometria desenvolve padrões e modelos matemáticos para medir estes processos, obtendo resultados para elaborar previsões gerenciais e tomada de decisão. Observa-se que a bibliometria, na forma atual, configura-se como uma valiosa ferramenta auxiliar no processo de gestão de informação.

Com relação às bases de dados bibliográficos, observa-se também uma diversificação na sua finalidade original, qual seja a busca e recuperação da informação. Atualmente, as bases de dados tornaram-se multifuncionais e constituem-se em importantes repositórios de matéria-prima intelectual para ser trabalhada desde que os dados estejam consistentes. E a consistência demanda avaliação.

Para a realização do estudo proposto – produção de indicadores da atividade de pós-graduação – foi inicialmente gerada uma nova base de dados constituída somente de dissertações e teses. Portanto, dos 11.600 itens totais da base *PTC*, foram selecionados 1.190 itens, identificados na base original pelo campo “Tipo de documento”.

A análise de consistência dessa nova base de dados, doravante denominada Base de Dissertações e Teses, foi efetuada com o uso dos *softwares* bibliométricos *DATAVIEW* e *INFOTRANS*, respectivamente. Trata-se de *softwares* de consistência para tratamento de dados bibliográficos.

O *DATAVIEW*, *software* bibliométrico elaborado pelo *Centre de Recherche Rétrospective de Marseille (CRRM) da Universidade Aix-Marseille III*, na França, é utilizado para a produção de listas de freqüência e matrizes de co-ocorrência sob a forma de tabelas e gráficos. Já o *INFOTRANS* é um *software* alemão da empresa *luK (Information und Kommunikation)* utilizado para a transformação dos registros bibliográficos em registros bibliométricos. Também 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 (KOBASHI e SANTOS, 2006).

Deve-se registrar que estes dois *softwares* foram utilizados na versão operacional *DOS*. Certamente existem versões mais modernas, porém o custo é inviável para aquisição por instituições públicas de pesquisa.

Por meio do *software DATAVIEW* foram emitidas listas, em ordem alfabética, dos conteúdos de cada campo selecionado para a análise: orientador, tipo de documento, local de defesa, ano de defesa e descritores de assunto. Nestas listas, comparando-se os itens de cada campo, foram detectadas as seguintes inconsistências:

- orientador: foram encontradas diferentes formas de entrada para o nome de um mesmo orientador;
- tipo de documento: diferentes formas de entrada foram utilizadas para identificar tanto as teses como as dissertações;
- local da defesa: foram detectadas diferentes formas de entrada para uma mesma instituição, inclusive no caso do próprio IPEN, ocasionadas tanto pelas alterações de nome que o instituto adotou no decorrer dos anos, quanto pela falta de padronização de entrada.

O campo “ano de publicação” não apresentou problemas; o campo “descritores” continha um número irrelevante de inconsistências. Os outros campos da base, como o número de localização no acervo ou dados complementares do documento, não foram analisados por não serem relevantes para os fins deste estudo.

As inconsistências detectadas causaram surpresa, pois a inserção de dados, em todas as bases da biblioteca do IPEN, é feita de forma criteriosa, atividade para a qual os técnicos que respondem pelas atividades de indexação, catalogação e registro, são treinados de forma cuidadosa.

Para a correção dos erros foi utilizado o *INFOTRANS*, o qual permite a padronização automática de dados, complementação de campos com novos dados e reorganização geral da base pela eliminação de registros duplicados. O *software* possibilita, ainda, a localização de uma seqüência de caracteres em um

determinado campo e sua substituição. Por exemplo, no campo “Tipo de documento” foram encontradas entradas como: Tese, Tese de doutoramento, Doutorado. Analisando-se tais entradas, optou-se pela adoção dos seguintes padrões para este campo: Tese (Doutoramento) e Dissertação (Mestrado). Assim, utilizando o *software* *INFOTRANS*, estes erros foram localizados e substituídos pelas entradas adotadas.

Os outros campos, onde também foram constatadas inconsistências, foram corrigidos da mesma forma: identificaram-se os erros por meio das listas emitidas pelo *DATAVIEW*, adotou-se um padrão de entrada e os dados incorretos foram substituídas pelos adotados, via *INFOTRANS*. Esta análise de consistência mostrou-se útil também para corrigir a base original *PTC*, possibilitando atingir também, uma melhoria na função de busca e recuperação de dados.

Para a análise temática, foi necessário criar novos campos: categoria de assunto e área de concentração do curso de pós-graduação. As categorias de assuntos foram identificadas por meio de pesquisa na base de dados do *INIS*. A busca foi efetuada pelo nome do autor e título do trabalho. As mesmas categorias atribuídas aos trabalhos na referida base foram adotadas na base de dados deste estudo; as áreas de concentração foram fornecidas pela comissão de pós-graduação, dessa forma, as informações coletadas para os novos campos foram inseridas em todos os registros da base de dados.

Com os dados preparados e “higienizados”, passou-se novamente à utilização do *DATAVIEW*, desta vez para a emissão de listas de freqüência para as análises bibliométricas. Foram elaboradas as seguintes listas: a) assuntos pesquisados; b) distribuição dos principais assuntos por quinquênios; c) orientadores dos trabalhos; d) descritores de assunto; e) áreas de concentração.

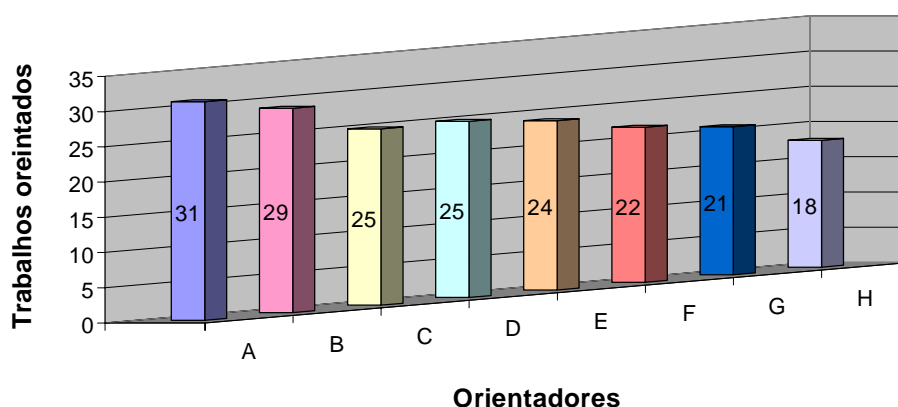
Essas listas foram, em seguida, exportadas para o *software* *EXCEL*, construindo-se os gráficos necessários para a melhor visualização dos resultados.

4. Resultados obtidos

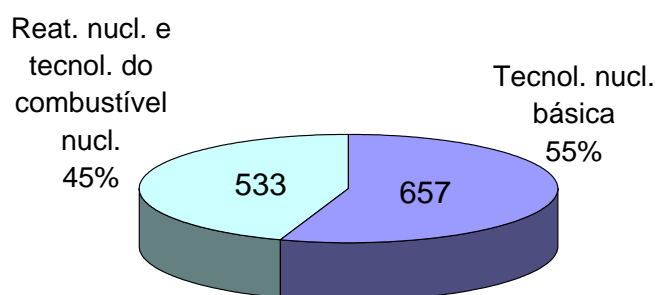
Cada campo verificado e preparado forneceu um tipo de informação. Neste trabalho são comentados, de forma concisa, alguns dos resultados obtidos. Mas,

com estes mesmos, dados é possível realizar inúmeras análises, desde as mais simples até análises mais elaboradas, correlacionando variáveis, como, por exemplo, co-ocorrência de palavras-chave, que poderão ser melhor visualizadas por meio de gráficos ou mapas da ciência.

O Gráfico 1, mostra os oito docentes que orientaram o maior número de trabalhos durante os trinta anos de existência do programa de PG do IPEN.

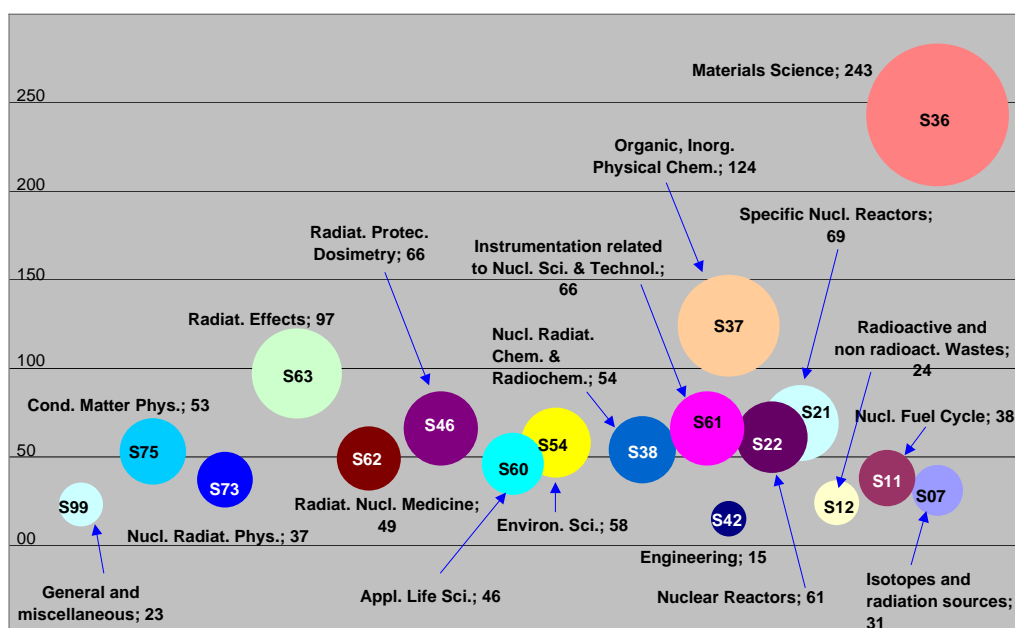


Graf. 1 - Docentes com maior número de orientações



Graf. 2 - Distribuição dos trabalhos por área de concentração

No Gráfico 2 observa-se o número de dissertações e teses distribuídas pelas áreas de concentração do programa de PG do IPEN. A partir de 1997 mudou a denominação destas áreas, manteve-se as anteriores, por ser considerado um dado irrelevante neste trabalho.



Graf. 3 - Ranking dos assuntos mais pesquisados

Para obter o Gráfico 3, (SANTOS et al, 2006) utilizou-se as informações do campo “categoria de assunto”. São apresentados os resultados da distribuição dos 18 temas mais pesquisados no programa de PG. O tamanho dos círculos corresponde ao número de teses e dissertações de cada categoria. As denominações das categorias foram mantidas em inglês, uma vez que foram retiradas da lista de categorias de assuntos do INIS (INIS, 2002).

Estes resultados demonstram, em nível de assunto, as características da produção científica na área nuclear nos últimos trinta anos, registradas nas dissertações e teses do curso de Pós-Graduação. Este tipo de informação auxilia a direção da instituição a identificar quais temas ou áreas de pesquisa estiveram em foco nos últimos trinta anos, quais foram, ou são as frentes de pesquisa. São informações que, correlacionadas aos cenários políticos, sociais e econômicos se transformam em informação estratégica para a tomada de decisão relativas às políticas de pesquisa.

5. Considerações finais

As bases de dados bibliográficos podem cumprir múltiplas funções. A função mais tradicional é a recuperação de informação. Nos últimos anos, essas bases têm sido crescentemente utilizadas para produzir indicadores

bibliométricos. Para isso, no entanto, os dados precisam ser consistentes, aptos a serem classificados e quantificados de modo coerente.

Observou-se no presente estudo, dedicado inicialmente à identificação do percurso de pesquisa do Programa de Pós-graduação do IPEN, que os métodos e *softwares* bibliométricos são bastante úteis para a análise de consistência das bases de dados. A identificação dos erros permite fazer correções nas bases de dados originais, de modo a promover melhor recuperação de informação. É fato conhecido que erros nos registros bibliográficos comprometem a qualidade da recuperação da informação e, por conseguinte, a qualidade dos serviços prestados pelas unidades de informação.

No entanto, é necessário enfatizar que a função dos *softwares* bibliométricos, como se procurou demonstrar, não é a de promover a recuperação de informação. Esses *softwares*, ao serem utilizados para fins gerenciais, mostram de forma quantificada e sistemática, por campo, os dados da base original. Pode-se, desse modo, detectar erros e inconsistências e corrigi-los.

A qualidade da recuperação da informação e dos indicadores de produção científica está largamente associada à qualidade das bases de dados. A qualidade destas últimas depende de processos rigorosos de tratamento da informação, para os quais concorrem as estruturas de metadados, as políticas de indexação e os vocabulários controlados. No entanto, é necessário manter políticas de gestão de qualidade das bases de dados, as quais necessitam de instrumentos que permitam “olhar” os dados de forma sistemática, promover sua “higienização” (correção de dados, eliminação de duplicações) e, desse modo, reorganizar as bases como um todo.

A consistência das bases de dados é, certamente, um trabalho que requer inversão de tempo e de recursos. No entanto, é um investimento que vale a pena se considerarmos que as bases de dados são dispositivos que permitem dar materialidade ao ciclo da informação, isto é, transformar dado em informação e informação em conhecimento (BARRETO, 1998).

Bibliografia

BARRETO, Aldo de A. A oferta e a demanda da informação: condições técnicas, econômicas e políticas. **Ciência da informação**, v. 27, n. 2, p. 122-127, 1998. Disponível em: <www.scielo.br>. Acesso em: 27 mar. 2007.

IGAMI, M.P.Z. **Análise da produção científica do Instituto de Pesquisas Energéticas e Nucleares, período, 1988-2003**. São Paulo: IPEN/CNEN-SP Divisão de Informação e Documentação Científicas, 2005. (trabalho de conclusão de disciplina na Pós-graduação – CBD 5819 – ECA USP).

INSTITUTO DE PESQUISAS ENERGETICAS E NUCLEARES. **Plano diretor, 2004**. São Paulo: Divisão de Planejamento e Inovação tecnológica, IPEN/CNEN-SP, 2000-2005.

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. 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., 2006.

PEREIRA, M. da N.F.; RIBEIRO, Claudio J.S.; TRACTENBERG, Leonel; MEDEIROS, Priscilla L. Bases de dados na economia do conhecimento: a questão da qualidade. **Ciência da informação**, v. 28, n. 2, p. 215-223. Disponível em: <www.scielo.br>. Acesso em: 26 mar. 2007.

ROSTAING, H. **La bibliométrie et ses techniques**. Toulouse: Co-édition Sicences de la Societé et CRRM, 1996.

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.

TAGUE-SUTCLIFFE, J. An introduction to informetrics. **Information Processing & Management**. v. 28, n. 1, p. 1-3, 1992.