

# A PRODUÇÃO CIENTÍFICA EM ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO NO BRASIL: uma análise bibliométrica do GT-2 da ANCIB

Ely Francina Tannuri de Oliveira\*  
Maria Cláudia Cabrini Gracio\*\*

## RESUMO

O objetivo desta pesquisa é analisar a produção científica da ANCIB, no grupo de trabalho GT-2, Organização e Representação do Conhecimento, no período de 2003 a 2008, através da Análise de Domínio, na abordagem que utiliza os indicadores bibliométricos, especificamente indicadores de produção e de ligação. *Os indicadores de produção* constituem-se na contagem do número de publicações de livros, artigos, publicações científicas, relatórios, por instituição, área de conhecimento ou país. Têm a finalidade de destacar e dar visibilidade à frente de pesquisa, de forma a evidenciar o referencial teórico-epistemológico dominante na área. *Os indicadores de ligação* são baseados na coocorrência de autoria, de citações e de palavras, usados para o mapeamento do conhecimento e construção da rede de relacionamento entre os pesquisadores, instituições ou países, e utilizam técnicas de análise estatística, matemática e computacional. O procedimento de pesquisa se deu a partir do levantamento das 134 pesquisas apresentadas no período, onde consta a referência do trabalho, com o resumo e palavras-chave correspondentes. Os dados foram lançados na base de dados *Access*, a partir da qual se construíram tabelas transportadas para o software *Excel*, para as variáveis: autores mais produtivos; temáticas mais frequentes dos autores mais produtivos retiradas através de palavras-chave; produtividade das instituições e coautoria institucional. Construiu-se a rede de colaboração científica entre as instituições, utilizando-se o software *Pajek*; utilizando-se o pacote *SPSS*, realizou-se a análise multivariada de *clusters*, através do método *Ward*. Os resultados apontam a existência de compatibilização entre autores mais produtivos, em suas instituições de origem e instituições mais produtivas, bem como as temáticas mais pesquisadas, que se apresentam, em geral, de forma panorâmica e dispersa. A rede de colaboração inter-institucional apresenta-se frágil e pouco densa, o que demonstra pouca colaboração entre as instituições, porém, em alguns casos, significativa colaboração intra-institucional.

**Palavras-chave:** Produção Científica. Análise Bibliométrica. Rede de Colaboração Científica. Frente de Pesquisa do GT-2.

## 1 INTRODUÇÃO

A ANCIB - Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação - é uma sociedade civil, sem fins lucrativos, fundada em junho de 1989, em decorrência dos esforços de alguns Cursos e Programas de Pós-Graduação da área, admitindo sócios institucionais (Programas de Pós-Graduação em Ciência da Informação) e sócios individuais (professores, pesquisadores, estudantes de pós-graduação e profissionais advindos dos programas). Sua finalidade é acompanhar e estimular as atividades de ensino de pós-

graduação e de pesquisa em Ciência da Informação, no Brasil. Desde sua criação, tem se projetado, no país e fora dele, como uma instância de representação científica e política importante para o debate das questões pertinentes à área de informação. As atividades da ANCIB estruturam-se em duas frentes: os Programas de Pós-Graduação *stricto sensu*, que são representados pelos seus coordenadores; e o ENANCIB, encontro nacional de pesquisa promovido pela ANCIB, que constitui grupos de debates e reflexões que reúnem pesquisadores e socializam a produção do conhecimento novo em temas especializados da Ciência da Informação, organizados em Grupos de Trabalho (GTs)

Os ENANCIBs vêm acontecendo desde 1994, propiciando ao longo desses anos edições de um fórum privilegiado para o debate da pesquisa conduzida pela área. Agrega pesquisadores e programas de pós-graduação brasileiros para um trabalho de prospecção e aprofundamento, que envolve desde o desenvolvimento de políticas até as investigações acerca dos fundamentos epistemológicos da área. Inicialmente concebidos para uma periodicidade bienal, os ENANCIBs passaram, a partir de 2005, a ter periodicidade anual, na expectativa de regularizar o fluxo e a circulação da crescente produção científica desenvolvida especialmente pela comunidade dos Programas de Pós-Graduação na área.

Através do amplo diálogo entre os pesquisadores, os ENANCIBs fazem avançar as atividades de geração de conhecimento, incentivando a reflexão sobre temas e tendências acadêmicas da atualidade. Hoje, a temática das discussões investigativas se encontra abrangida por nove grupos de trabalhos, a saber: GT-1: Estudos Históricos e Epistemológicos da Informação; GT-2: Organização e Representação do conhecimento; GT-3: Mediação, Circulação e Uso da Informação; GT-4: Gestão da Informação e do Conhecimento nas Organizações; GT-5: Política e Economia da Informação; GT-6: Informação, Educação e Trabalho; GT-7: Produção e Comunicação da Informação em CT&I; GT-8: Informação e Tecnologia; e GT-9: Museologia, patrimônio e informação.

Considerando a importância das atividades dos GTs para o desenvolvimento científico da Ciência da Informação no Brasil, tem-se por proposição analisar a produção científica do GT2 - Organização e Representação do conhecimento – da ANCIB.

\* Ely Francina Tannuri de Oliveira PPGCI - Universidade Estadual Paulista-Unesp/Marília. Profa. Assistente Doutora do Departamento de Ciência da Informação. E-mail: [etannuri@flash.tv.br](mailto:etannuri@flash.tv.br)

\*\* Maria Cláudia Cabrini Gracio. PPGFil - Universidade Estadual Paulista-Unesp/Marília. Profa. Assistente Doutora. E-mail: [cabrini@marilia.unesp.br](mailto:cabrini@marilia.unesp.br)

Para tanto, observa-se a seguinte ementa do GT em questão:

Teorias, metodologias e práticas relacionadas à organização e preservação de documentos e da informação, enquanto conhecimento registrado e socializado, em ambiências informacionais tais como: arquivos, museus, bibliotecas e congêneres. Compreende, também, os estudos relacionados aos processos, produtos e instrumentos de representação do conhecimento (aqui incluindo o uso das tecnologias da informação) e as relações inter e transdisciplinares neles verificadas, além de aspectos relacionados às políticas de organização e preservação da memória institucional. (<http://www.ancib.org.br/grupos-de-trabalho/gt-2-organizacao-e-representacao-do-conhecimento/>)

A partir desta ementa, observa-se que os temas de pesquisa abarcados pelo GT-2 inserem-se em um espectro bastante amplo, em cujo âmbito podem-se destacar os seguintes eixos temáticos relativos à organização e representação do conhecimento: questões epistemológicas (teorias, metodologias); ambiências; processos; produtos; instrumentos; interdisciplinaridades; políticas; memória. (GUIMARÃES; LIBERATORE, 2004).

Dessa forma, a presente pesquisa objetiva, de forma geral, analisar a produção científica no GT-2 nos últimos cinco ENANCIBs, por meio da Análise de Domínio, na abordagem que utiliza os indicadores bibliométricos, especificamente os indicadores de produção e ligação, de modo a mostrar a frente de pesquisa no grupo Organização e Representação do Conhecimento - GT-2, no Brasil.

De forma mais específica, busca identificar os pesquisadores mais produtivos, as temáticas mais freqüentes, as instituições mais produtivas e a rede de colaboração determinada pelas co-autorias institucionais.

A escolha pelo GT-2 da ANCIB como universo de pesquisa deveu-se, primeiramente, à carência de estudos relativos à análise e avaliação da produção científica dos ENANCIBs, e especialmente por ser uma subárea de pesquisa que tem apresentado aumento significativo de pesquisas nos diferentes anos, a saber: em 2003 (V ENANCIB - Belo Horizonte - MG), 17 trabalhos; em 2005 (VI ENANCIB - Florianópolis-SC), 26 trabalhos; em 2006 (VII ENANCIB - Marília-SP), 23 trabalhos; e em 2007 (VIII ENANCIB - Salvador-BA), 45 trabalhos; e em 2008 (IX ENANCIB - São Paulo-SP), 23 trabalhos, totalizando 134 trabalhos.

Justifica-se ainda a análise dos trabalhos dos cinco últimos ENANCIBs pelo fato de, somente a partir de 2003, no V ENANCIB em Belo Horizonte- MG, os anais passarem a ser disponibilizados em suporte digital (CD-ROM).

Destaque-se ainda que, a partir do momento que avalia sua produção, um grupo pode alcançar parâmetros que tornam possível reavaliar e repensar seus objetivos, reprogramando

suas estratégias de crescimento e necessidades. Além disso, quando o volume da produção científica passa a ser visível nas instituições ou grupos, se faz necessária a utilização de metodologias para avaliá-la.

## **2 ANÁLISE DA PRODUÇÃO CIENTÍFICA ATRAVÉS DE ANÁLISE BIBLIOMÉTRICA**

A *produção científica*, entendida como o conjunto de publicações geradas durante a realização e após o término de pesquisas, vem sendo cada vez mais estudada, especialmente nos últimos 40 anos, quando a explosão documentária impôs, de certa forma, a necessidade da criação de instrumentos para avaliação da ciência, particularmente de instituições de pesquisas e pesquisadores, especialmente em países, como o Brasil, onde o desenvolvimento da ciência encontra-se mais diretamente relacionado ao sistema de educação superior.

Entretanto, os estudos de produção científica enfrentam desafios, considerando que esta integra um grande sistema social, a ciência, cujas funções consistem em disseminar conhecimentos, assegurar a preservação de padrões e atribuir crédito e reconhecimento aos autores de trabalhos de relevante contribuição para o desenvolvimento das idéias em diferentes campos. (MACIAS-CHAPULA, 1998, p.136)

Para se avaliar questões de produtividade científica, a *Análise de Domínio* (A.D.) aparece hoje, notadamente no universo da Ciência da Informação, como um dos principais respaldos teóricos. Utilizada pela primeira vez em 1980, na área de *Ciência da Computação*, por Neighbors, como “uma tentativa de identificar objetos, operações e relações entre o que peritos de determinado domínio percebem como importante” (KERR, 2003, p.11), a análise de domínio foi pela primeira vez utilizada na Ciência da Informação por Hjørland e Albrechtsen, no início da década de 90 (HJØRLAND, 2002b, p.259).

O *Domínio* pode ser compreendido como “uma área de especialidade, um conjunto literário ou um grupo pessoas trabalhando juntas em uma organização” (MAI, 2005, p.605). Sob o ponto de vista desta pesquisa, é uma área de conhecimento, atividade, interesse, onde é demarcado um determinado conhecimento com limites definidos e onde profissionais ou grupos estão articulados tanto em pensamento como em linguagem.

Conceito bastante amplo, como também controverso, a *Análise de Domínio* aplicável a diferentes áreas do conhecimento é um processo através do qual se pode sinalizar o que é mais importante ou significativo dentro de algum campo ou contexto científico, de modo a se analisar tendências, relações existentes ou processos. Em suma, pode-se dizer que a Análise

de Domínio constitui um dos instrumentos metodológicos que permitem analisar o *comportamento da ciência* em um dado campo.

Como destaca Hjørland (2002a), diversificadas são as abordagens de A.D., tais como: a produção de guias de literatura, a elaboração de classificações especiais e tesouros, a indexação e recuperação da informação, os estudos empíricos de usuários, os *estudos bibliométricos*, os estudos históricos, os estudos de documentos e gêneros, os estudos epistemológicos e críticos, os estudos terminológicos, as estruturas de instituições da comunicação científica, cognição, conhecimento e inteligência artificial.

Especificamente o âmbito dos estudos bibliométricos tem sido uma fértil seara para a análise de um domínio - área de especialidade, grupo ou diferentes conjuntos – na medida em que analisa sua produção científica, por meio de indicadores quantitativos. Têm suas origens na década de 30, quando Bradford enunciou a *Lei da Dispersão Bibliográfica*. Os estudos bibliométricos vêm ganhando cada vez mais espaço, tanto no universo acadêmico quanto em órgãos governamentais e órgãos multinacionais, para fins de direcionamento de recursos para a pesquisa e desenvolvimento.

As avaliações quantitativas por meio de indicadores bibliométricos apontam para esse ou aquele cientista ou instituição, essa ou aquela área do conhecimento, esse ou aquele país, como mais férteis ou mais produtivos. Nesse contexto, a análise da produção científica de um país, de uma região ou instituição científica envolve um conjunto expressivo de indicadores bibliométricos, que se agrupam em *indicadores de produção*, *indicadores de ligação e indicadores de citação* (OKUBO,1997; SPINAK, 1998; NARIN et al., 1994; CALLON et al.,1993).

*Os indicadores de produção* constituem-se na contagem do número de publicações de livros, artigos, publicações científicas, relatórios, por instituição, área de conhecimento ou país, bem como indicadores percentuais, taxas de crescimento, distribuição da produtividade de autores (Lei de Lotka), distribuição do uso de vocabulário (distribuição de Zipf), distribuição de produtividade de periódicos (Lei de Bradford), entre outros. Esses indicadores de produção revelam pesquisadores, instituições ou mesmo periódicos mais produtivos em determinada área, grupo ou país, com a finalidade de destacar e dar visibilidade a uma determinada frente de pesquisa, de forma a evidenciar o referencial teórico-epistemológico dominante na área.

*Os indicadores de ligação* são baseados na coocorrência de autoria, de citações e de palavras, e se utilizam, para o mapeamento do conhecimento e construção da rede de relacionamento entre os pesquisadores, instituições ou países, valendo-se de técnicas de

análise estatística, matemática e computacional. Nesse contexto, a análise de coautoria ocupa espaço significativo, sendo medida pelo número de publicações em colaboração de autores ou instituições e empregada para identificar e mapear a cooperação regional, nacional ou internacional.

Os *indicadores de citação* baseiam-se na medida do número de citações, fator de impacto (FI) e outros índices relevantes, tendo como premissa básica de medida do desempenho científico o fato de uma publicação necessitar de certa qualidade básica para que possa efetivamente gerar impacto em uma comunidade científica. Desse modo, pode-se dizer que o número de citações que uma dada publicação recebe revela seu grau de permeabilidade na comunidade científica a que se destina e seu impacto no universo de conhecimento dos demais pesquisadores.

O pesquisador Smith, em 1958, citado por Balancieri et al (2005, p.3), foi um dos primeiros a observar o crescimento da publicação em coautoria e a sugerir o seu uso como uma medida de colaboração entre grupos de pesquisadores ou instituições de um mesmo país ou a colaboração internacional.

Na década de 1960, destaca-se o início dos estudos na área da colaboração científica. Entre as constatações nesse período, observou-se que a colaboração científica se inicia nas relações entre orientador e orientando e, especialmente, no âmbito dos “colégios invisíveis”.

Assim, a colaboração científica é um dos fenômenos mais visíveis observados na construção da ciência ao longo da história. Price (1963), citado por Barragan et al. (2006, p.414), foi um dos primeiros pesquisadores a apontar as evidências empíricas do aumento das coautorias na ciência.

Segundo Balancieri et al (2005, p.2), “a colaboração científica oferece uma fonte de apoio para melhorar o resultado e maximizar o potencial da produção científica”, que se expressa por meio de uma rede onde os diferentes colaboradores se relacionam.

Segundo Wasserman e Faust (1994, p.9), “o termo ‘rede social’ se refere ao conjunto de atores e suas ligações entre eles”. A análise de rede tem por objetivo modelar as conexões entre os atores, a fim de retratar, descrever e representar a estrutura de um grupo, quer seja composto por países, instituições ou pessoas.

O conceito de análise de rede social desenvolveu-se a partir de uma confluência da teoria social com a metodologia matemática, estatística e computacional.

A rede de colaboração científica entre autores ou instituições (OLMEDA GÓMEZ; PERIANEZ-RODRIGUEZ; OVALLE-PERANDONES, 2008) supõe um compartilhamento de ideais centrais de um projeto, os objetivos e as consequências que geraram essas idéias: é

necessário um clima de confiança, o estabelecimento de uma divisão de trabalho e interação entre os investigadores, compartilhamento de informações e coordenação destas diferentes relações do investimento conjunto e redação final. Assim, a coautoria reflete o rol possível de intercâmbios e trocas entre os pesquisadores, como são, por exemplo, as conversas informais e as discussões em congressos e palestras advindas das apresentações de pesquisas.

O importante ganho resultante das autorias múltiplas, sejam duplas, triplas ou *n-uplas*, quando comparado ao trabalho de pesquisadores isolados, é a ampliação do repertório de abordagens e ferramentas, que advém do intercâmbio de informações e da produtividade que se verifica quando grupos, pesquisadores ou instituições distintas juntam esforços no sentido de determinada meta, promovendo a interação entre os pesquisadores.

### **3 METODOLOGIA**

O levantamento dos dados se deu a partir de súmula constituída dos 134 pesquisas apresentadas, onde constou a referência do trabalho, palavras-chave correspondentes e texto. Esse universo foi constituído de 17 trabalhos apresentados no V ENANCIB, que ocorreu em 2003 (Belo Horizonte); dos 26 apresentados em 2005, no VI ENANCIB (Florianópolis); 23 trabalhos no VII ENANCIB, em 2006 (Marília); 45 trabalhos em 2007 (Salvador); e 23 apresentados em 2008 (São Paulo).

Os dados foram lançados na base de dados *ACCESS*, tomando-se como variáveis iniciais, para alimentação da mesma, os autores, instituições de origem, palavras-chave e autor por palavras-chave. A partir destes arquivos de dados, cruzaram-se as variáveis, constituindo-se tabelas transportadas para o software *Excel*, a saber: autores mais produtivos; autor por palavras-chave; produtividade das instituições; tipo de autoria (simples ou co-autoria); frequência de coautoria institucional.

As dificuldades encontradas na alimentação da base e posterior transporte de dados para o software *Excel* foram relativas aos registros feitos, nem sempre com a devida precisão, especialmente em relação aos nomes das instituições de origem, autores e palavras-chave, estas últimas, nos dois primeiros ENANCIBs nem sempre definidas, exigindo-se, algumas vezes, a leitura dos textos para sua geração. Alguns ajustes e revisões na base se tornaram necessários também na fase de análise de dados. Porém, a maior dificuldade encontrada na alimentação dos dados foi relativa à instituição de origem dos pesquisadores, em virtude da duplicidade de registros, ora por programa de pós-graduação, ora por vínculo empregatício,

fato que exigiu uma retomada dos registros, com o objetivo de torná-los compatíveis e uniformes. Deste modo, quando o autor apresentou dupla filiação institucional, uma por vínculo empregatício e outra por participação em Programas de Pós-Graduação, seja como aluno ou docente, optou-se, no caso de o vínculo empregatício ser professor/pesquisador, pela instituição de pesquisa onde exercia a docência-pesquisa. Nos demais casos, quando o autor não é docente, mas apresenta vínculo empregatício, registrou-se o programa de Pós-Graduação como filiação do mesmo.

Ainda em relação ao registro das instituições de origem dos pesquisadores, decidiu-se que a produção científica decorrente do Programa de Pós-Graduação do IBICT no decorrer dos dois convênios historicamente observados - IBICT-UFRJ e IBICT-UFF - foram computados como IBICT na medida em que esse órgão é o responsável pela concepção do programa original e pela base do corpo de pesquisadores que sustenta essa proposta. No entanto, computou-se para a UFF ou a UFRJ a produção científica decorrente de seu corpo específico de professores.

Tanto na situação citada, como em outras nas quais houve dúvida em relação à filiação institucional do autor, recorreu-se à Plataforma Lattes a fim de se tomar uma decisão consistente.

Um total de 174 pesquisadores participou no período analisado e foram considerados pesquisadores mais produtivos aqueles que fizeram pelo menos três produções no GT-2. Considerando a apresentação de pelo menos um trabalho por ano por pesquisador, o critério se justifica pela hipótese de o mesmo ter participado de mais da metade dos encontros, presença considerada significativa e visível.

Buscou-se estabelecer os temas mais estudados pelos pesquisadores mais produtivos a partir da relação entre eles e as palavras-chave mais citadas. Para cada autor, retirou-se o conjunto de palavras-chave, a partir do conjunto de seus trabalhos, gerando os temas mais candentes por autor e classificados de acordo com a ementa do próprio GT-2.

As temáticas mais relevantes, considerando-se a totalidade das 134 pesquisas, foram retiradas a partir das palavras-chave mais frequentes. Partiu-se do pressuposto de que tais palavras determinam as questões mais estudadas nos ENANCIBS e, portanto, aquelas que mais se constituem em foco de pesquisas no GT2.

Quanto às instituições, 44 foram cadastradas, destacando-se como as mais produtivas aquelas que apresentaram pelo menos 3 trabalhos, pela mesma justificativa considerada em relação ao corte efetuado para os autores em relação ao número de trabalhos apresentados.

Quanto às coautorias, separaram-se, em um primeiro momento, os trabalhos de autoria simples dos trabalhos de autoria dupla, tripla, quádrupla ou *n-úpla*, verificando-se que, dos 134 trabalhos apresentados, 43 são de autoria simples e 91 em coautorias. A seguir, montou-se a matriz de coautorias institucionais, com as 39 instituições que apresentaram os 91 trabalhos em colaboração. Construiu-se a rede de colaboração científica institucional, utilizando-se o software *Pajek*. Com a finalidade de mostrar as afinidades, similaridades e dissociações em relação ao comportamento das colaborações científicas entre instituições, utilizando o pacote *SPSS*, realizou-se a análise multivariada de *clusters*, através do método *Ward*, com variáveis padronizadas. A visualização da aglomeração dos *clusters* foi feita através do dendograma.

#### 4 APRESENTAÇÃO E ANÁLISE DOS DADOS

A partir da organização dos dados relativos à autoria, registrou-se um total de 272 autorias realizadas por 174 autores nos 134 trabalhos, o que indica uma média de aproximadamente 1,6 autorias por autor e, aproximadamente, em média, 2 autorias por trabalho. Deste total de 174 pesquisadores, 24 apresentaram 3 ou mais trabalhos no período sob análise (14% do total), considerando-se este o conjunto dos autores mais produtivos, apresentados nos dados da Tabela 1, a seguir.

Tabela 1 - Autores mais produtivos<sup>1</sup>

| <b>Autor</b>                         | <b>Total de Publicações</b> |
|--------------------------------------|-----------------------------|
| <b>CAMPOS, M.L.A. (UFF)</b>          | 9                           |
| <b>FUJITA, M.S.L. (UNESP)</b>        | 8                           |
| <b>GUIMARÃES, J.A.C. (UNESP)</b>     | 7                           |
| LIMA, G.A.B. (UFMG)                  | 7                           |
| <b>SOUZA, R.F. (IBICT)</b>           | 5                           |
| BRASCHER, M. (UNB)                   | 4                           |
| <b>DODEBEI, V.L (UNIRIO)</b>         | 4                           |
| <b>LARA, M.L.G. de (USP-ECA)</b>     | 4                           |
| <b>MARCONDES, C.H. (UFF)</b>         | 4                           |
| MOREIRA, M. P. (PUC-MG)              | 4                           |
| <b>MURGUIA, E.I. (UNESP/USP-RIB)</b> | 4                           |
| SOUZA, R.R. (UFMG)                   | 4                           |
| <b>TÁLAMO, M.F. (USP/PUC-CAMP.)</b>  | 4                           |
| CAFE, L.M.A (UFSC)                   | 3                           |
| CERÁVOLO, S.M. (UFBA)                | 3                           |
| CORDEIRO, R.I.N. (UFF)               | 3                           |
| GALVÃO, M.C.B. (USP-RP)              | 3                           |
| GOMES, H.E. (CONS. INDEP.)           | 3                           |
| LIMA, M.L. (UNESP)                   | 3                           |

|                               |   |
|-------------------------------|---|
| MANINI, M.P. (UNB)            | 3 |
| MIRANDA, M. (UNIRIO)          | 3 |
| <b>MORAES, J.B.E. (UNESP)</b> | 3 |
| <b>MOURA, M.A. (UFMG)</b>     | 3 |
| <b>ODDONE, N. (UFBA)</b>      | 3 |

<sup>1</sup> Os autores em negrito são bolsistas de produtividade em pesquisa do CNPq

Observa-se, nesse universo apresentado, que 20 autores provêm de instituições que possuem programas de pós-graduação na área, revelando que a frente de pesquisa da área se encontra predominantemente locada (83%) em espaços formais de produção do conhecimento, com um índice mínimo de dispersão de pesquisadores. Neste grupo, observa-se uma produção relativamente equilibrada, com exceção dos 4 primeiros autores que respondem, juntos, por aproximadamente 22% do total dos trabalhos.

Outro aspecto a ser observado: 11 autores (46%) são bolsistas de produtividade em pesquisa do CNPq, aspecto que se mostra ainda mais significativo se considerarmos que a área de Ciência da Informação como um todo dispõe, hoje, de 37 bolsistas de Produtividade em Pesquisa. Isso mostra que a área se encontra fortemente inserida nos universos formais de apoio à pesquisa e que esses pesquisadores têm efetivamente buscado socializar e solidificar o conhecimento por eles produzido com apoio governamental.

Relativamente às temáticas dos autores mais produtivos, observa-se que as duas vertentes que integram a própria denominação do GT-2, a Organização e Representação da Informação e do Conhecimento, são as mais citadas nas palavras-chave com 26 e 17 incidências, respectivamente, e por um amplo espectro de autores. Em que pese a generalidade dessas temáticas, tal aspecto poder-se-ia explicar em virtude da necessidade de os autores tornarem efetivamente compreensível - e visível - a inserção temática de seu trabalho no âmbito do GT-2. A isto se alia a Recuperação da Informação (com nove incidências, em nove autores), pois a recuperação é o objetivo mediato de toda a área abrangida pelo GT-2.

Ainda, analisando as demais palavras-chave utilizadas, em ordem decrescente de incidência, observa-se uma forte preponderância das temáticas e questões ligadas aos

instrumentos da área, na medida em que termos como *ontologias*, *linguagens documentárias*, e *tesauros* ocupam um espaço significativo nos trabalhos. Um segundo grupo a destacar refere-se aos processos específicos da área, notadamente por meio de termos como *indexação* e *análise documentária*, em cujo âmbito se inserem questões como a *tematicidade*. Um terceiro grupo refere-se às interfaces da área, sejam elas devido a aportes interdisciplinares (*Terminologia*), sejam em virtude das ambiências informacionais (*Museologia*, *Arquivologia*, notadamente nos *arquivos fotográficos*), enquanto espaços de *memória*. Tangenciando a área, observa-se o impacto das tecnologias, em temas como o *patrimônio digital*, a *publicação eletrônica* e a *web semântica*.

Uma análise dos quatro primeiros autores mais produtivos (Campos, Fujita, Guimarães e Lima) revela um ponto de convergência nas duas vertentes temáticas básicas do GT-2: a da Organização e Representação do Conhecimento/Informação, reiterando o que ocorre na produção geral, como já comentado.

No entanto, em temáticas específicas, observam-se poucos temas de convergência, a exceção das linguagens documentárias ou de indexação, notadamente de Campos, Fujita e Guimarães. Nesse sentido, observa-se em Campos e Guimarães uma preocupação comum com temas como os sistemas de classificação e os tesauros.

A questão de aportes disciplinares na área é objeto de abordagem de Fujita, Lima e Guimarães, as duas primeiras mais voltadas para o universo das Ciências da Linguagem (Linguística e Terminologia), e o último com ênfase na Diplomática.

Provenientes da mesma instituição, Fujita e Guimarães apresentam produção voltada para a Formação e Atuação Profissional em Organização e Representação do Conhecimento e da Informação, a primeira mais voltada para a formação do indexador e o segundo, na dimensão da atuação do indexador, na abordagem ético-profissional.

Completando este quadro, observa-se que, enquanto Campos tem sua abordagem mais voltada para o universo teórico da área, Lima centra sua ênfase na dimensão aplicada, notadamente no que tange à indexação automática.

A seguir, apresenta-se o rol das 14 instituições que foram autoras de pelo menos três trabalhos no universo estudado, representando 32% do total das 44 instituições presentes.

Tabela 2 - Instituições mais produtivas<sup>1</sup>

| <b>Instituições</b> | <b>Número de Trabalhos</b> |
|---------------------|----------------------------|
| <b>UFMG</b>         | 26                         |
| <b>UNESP</b>        | 26                         |
| <b>UFF</b>          | 17                         |
| <b>UNB</b>          | 11                         |
| <b>IBICT</b>        | 10                         |
| <b>USP-ECA</b>      | 9                          |
| UNIRIO              | 8                          |
| UFBA                | 7                          |
| PUC-CAMPINAS        | 6                          |
| PUC- MINAS GERAIS   | 5                          |
| UFPB                | 4                          |
| UFSC                | 4                          |
| USP-RIBEIRÃO PRETO  | 4                          |
| UFC                 | 3                          |

<sup>1</sup> Instituições que possuem Organização e Representação da Informação como conteúdo de uma de suas linhas de pesquisa no Programa de Pós-graduação.

Em termos institucionais, observa-se que, dos 134 trabalhos analisados, 123 (92%) têm pelo menos uma das 14 instituições, listadas na Tabela 2, como autora do trabalho. Conseqüentemente, 32% das instituições são responsáveis por 92% dos trabalhos apresentados. Assim, por outro lado, apenas 11 (8%) não possuem como autora nenhuma das instituições listadas como as mais produtivas. Estes percentuais ratificam que a frente de pesquisa institucional em Organização e Representação do Conhecimento é realmente constituída por essas instituições.

Observa-se ainda que as seis instituições mais produtivas (UFMG, UNESP, UFF, UNB, IBICT e USP-ECA) constituem exatamente aquelas que possuem a Organização e Representação da Informação como conteúdo de uma de suas linhas de pesquisa, produzindo um total de 94 trabalhos, que corresponde a 70% do total dos 134 trabalhos. Esse aspecto configura-se igualmente significativo, na medida em que revela que a produção no tema em questão está centrada em instituições que apresentam esta temática contemplada de forma explícita entre suas linhas de pesquisa, e acaba por alavancar a produção científica na área.

A seguir, apresenta-se a rede de colaboração institucional construída com as 39 instituições dos trabalhos em colaboração. Destaca-se que as áreas dos círculos são proporcionais à frequência de coautorias dentro da própria instituição, a espessura dos



A UFMG destaca-se como uma das duas instituições mais produtivas, centrando uma segunda sub-rede significativa, composta por sete instituições, com forte relação com a PUC-MG, a segunda maior da rede, e a UFRRJ e alta frequência de coautoria interna.

Uma terceira sub-rede, composta por cinco instituições, está centrada na UFPB, caracterizando-se como a única sub-rede na qual não se observa colaboração com as demais instituições mais produtivas apontadas na Tabela 2.

A UNB centra uma pequena sub-rede composta por quatro instituições.

Uma última sub-rede composta por três instituições está centrada na UNESP, também observada como uma das duas mais produtivas e com a maior frequência de colaboração interna de toda a rede.

Observe-se que cinco instituições só apresentaram colaboração interna, aspecto que revela, por um lado, apenas interlocução entre os integrantes daquelas instituições e, por outro, ausência de dialogicidade nesse contexto científico, no tema. Dessas cinco instituições, quatro ainda não dispõem de programa de pós-graduação na área e a outra constitui uma instituição estrangeira, com mestrado e doutorado na área, que seguramente apresenta dialogicidade científica com instituições outras, no exterior.

Calculou-se a densidade total da rede de colaboração, tomando-se o quociente entre o número de conexões existentes (55) pelo total de conexões possíveis (741), resultando em 0,07, configurando 7% das possibilidades de conexões, indicador que sugere uma frágil relação entre as instituições.

A seguir apresenta-se o dendograma referente à classificação das 39 instituições, segundo as coautorias observadas.

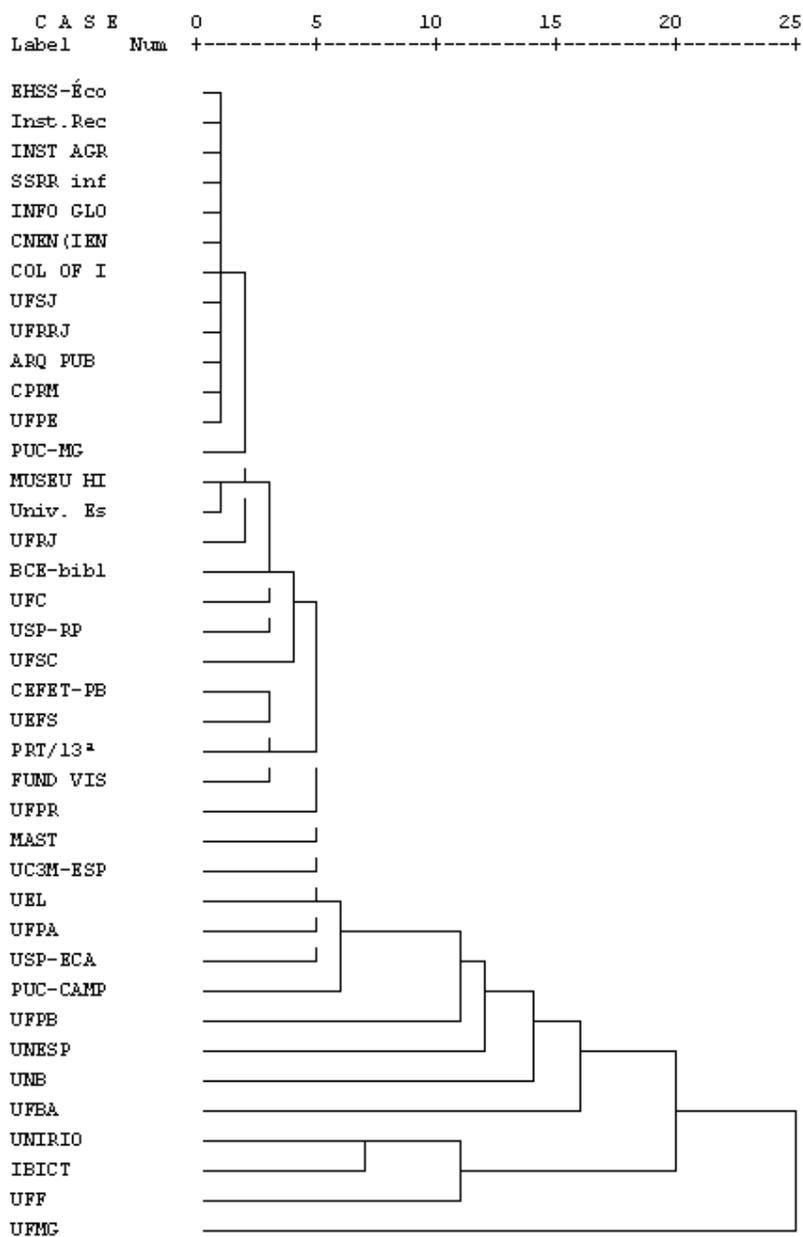


Figura 2. *Clusters* relativo à classificação das instituições, segundo coautorias

Na leitura do dendograma, destacam-se três grupos constituídos segundo o padrão de similaridade de colaboração institucional. O primeiro e maior, constituído por 35 instituições que se inicia com o EHSS-*École des Hautes Études en Sciences Sociales* até a UFBA, apresenta como característica principal o fato de serem instituições com frágil colaboração inter-institucional. O segundo *cluster* é constituído pela UNIRIO, pelo IBICT e pela UFF, pode ser caracterizado pelo fato de serem instituições com colaboração interna pouco

significativa, porém com significativo número de conexões com outros pares. O terceiro *cluster* contém uma única instituição, a UFMG, por apresentar comportamento diferenciado em relação aos outros dois *clusters*, a saber: alta colaboração interna e alto número de conexões com os pares.

## **5 CONSIDERAÇÕES FINAIS E RECOMENDAÇÕES**

Os dados coletados apontam um número significativo de pesquisadores e instituições trabalhando as questões de Organização e Representação do Conhecimento no Brasil, tema nuclear na área de Ciência da Informação.

Observa-se uma alta coesão entre autores mais produtivos e instituições com destaque na produção do conhecimento. A maioria dos autores mais produtivos está credenciada em Programas de Pós-graduação e participa de grupos de pesquisas. Além disso, muitos são bolsistas de produtividade em pesquisa do CNPq.

Em relação às instituições mais produtivas, destaca-se especialmente a região sudeste, onde se concentra o maior número de Programas de Pós-Graduação na área. Das 14 instituições mais produtivas, 6 possuem o Programa com Mestrado e Doutorado.

As temáticas mais frequentes contempladas nos trabalhos ocorrem, em geral, de forma bastante dispersa dentro da Organização e Representação do Conhecimento/Informação e de forma mais específica em linguagens documentárias, tesouros, indexação e análise documentária, terminologia, museologia e arquivologia (memória e arquivos fotográficos), bem como patrimônio digital, publicações eletrônicas e a web semântica.

A rede de colaboração institucional se apresenta, por um lado, com algumas instituições isoladas somente com coautoria intra-grupo e, por outro, instituições que apresentam a configuração de uma rede científica pouca densa, mas com sub-redes já formadas. Observa-se nesta rede uma escassa conexão entre as instituições mais produtivas, em geral, constituindo núcleo das cinco sub-redes identificadas. As instituições trabalham, de modo geral, com poucas parcerias, sem que se verifique a necessária dialogicidade acadêmica entre os programas de pós-graduação da área ou mesmo com instituições do exterior, aspectos que constituiriam importante recurso para ampliar a excelência acadêmica da área.

No tocante ao agrupamento das instituições por similaridade em relação ao comportamento de colaboração científica, destacaram-se 3 *clusters* com as seguintes características: um primeiro grupo, caracterizado pela baixa conexão inter-institucional; o

segundo, pela colaboração interna pouco significativa e destacada conexão inter-institucional; e o terceiro (unitário), que apresenta alta colaboração inter e intra-grupo.

Sugere-se que as questões de normalização, em relação especialmente às instituições de origem dos pesquisadores, sejam mais cuidadosamente tratadas, no sentido de facilitar os registros das mesmas em estudos similares.

Recomendam-se pesquisas ulteriores que contemplem o estudo e relação entre outras variáveis e análise de outros indicadores bibliométricos, tais como de citação e de ligação, neste mesmo GT, em outros GTs, bem como se estenda a outras instituições e ambientes.

Finalizando, acredita-se que o uso de dados coletados junto aos principais eventos da área, tal como os ENANCIBs, possa subsidiar a avaliação de políticas para as instituições de fomento à pesquisa, bem como a visualização de frentes de pesquisa nas diferentes sub-temáticas. Os dados da presente pesquisa, embora limitados em seu alcance pelo recorte no tempo, permitem uma apreciação do panorama da consistência da pesquisa no GT-2, constituindo-se em um primeiro olhar sobre a análise de sua produção científica.

**SCIENTIFIC PRODUCTION WITH REGARD TO KNOWLEDGE  
ORGANIZATION AND REPRESENTATION IN BRAZIL: bibliometrical analysis  
of ANCIB GT-2**

**ABSTRACT**

The aim of this research is to analyze ANCIB GT-2 working group, Knowledge Organization and Representation, covering the period from 2003 to 2008, through Domain Analysis, by means of an approach that uses bibliometrical indicators, specially the ones of production and linkage. *Production indicators* are those related to the counting of the amount of publications, reports for each institution, area of knowledge or country. Their purpose is to give emphasis and visibility to the research front so that the mainstream of theoretical-epistemological reference in the area becomes clear. *Linkage indicators* are based in the co-occurrence of authorship, citations and words used to map knowledge and build the networks among researchers, institutions or countries, and they use techniques of statistical, mathematical and computational analysis. Investigation started from a previous electronic search of 134 researches of *Access* databases, in which reference of the publication was indicated, with abstract and corresponding keywords. From that, tables were built and conveyed to *Excel* software for the variables: the most productive authors; the most frequent subjects of the most productive authors, identified through keywords; institutions productivity and institutional co-authorship. Collaboration network was built among institutions, using *Pajek* software; by using *SPSS*, *clusters* multi-varied analysis was done, by means of *Ward* method. Outcomes show compatibility among the most productive authors, in their institutions of origin and the most productive institutions, as well as among the most investigated subjects, that are overall

presented in a panoramic and scattering way. Inter-institutional collaboration network appears to be weak and lacking density, which demonstrates little collaboration among institutions, however, in some cases significant intra-institutional collaboration.

**Keywords:** Scientific production. Bibliometrical analysis. Scientific collaboration network. GT-2 front research.

### **Agradecimentos**

Agradecemos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio ao projeto aprovado no Processo 400237/2009-1, do qual esta pesquisa faz parte, constituindo-se em um recorte do mesmo.

Agradecemos também ao Prof.Dr. José Augusto Chaves Guimarães, da UNESP-Marília, pela valiosa colaboração na análise qualitativa dos dados.

### **REFERÊNCIAS**

ANCIB.Ementa do GT-2. Disponível em: <<http://www.ancib.org.br/grupos-de-trabalho/gt-2-organizacao-e-representacao-do-conhecimento/>>. Acesso em: 30 junho de 2009

BALANCIERI, R. et al. A análise de redes de colaboração científica sob as novas tecnologias da informação e comunicação: um estudo na Plataforma Lattes. **Ciência da Informação**, v. 34, n. 1, p. 64-77, 2005.

BARRAGAN, M. J.; GUERRERO-BOTE, V. P.; MOYA-ANÉGON, F. Colaboración científica de España com América Latina y el Caribe. In: ENCUESTRO ASOCIACIÓN DE EDUCADORES E INVESTIGADORES DE BIBLIOTECOLOGÍA, ARCHIVOLOGÍA, CIENCIAS DE LA INFORMACIÓN Y DOCUMENTACIÓN DE IBEROAMÉRICA Y EL CARIBE, 7., 2006, Marília. **Anales ... Marília**, 2006. Disponível em: <[http://www.edibcic.org/VII\\_EDIBCIC/VIIEDIBCIC.htm](http://www.edibcic.org/VII_EDIBCIC/VIIEDIBCIC.htm)>. Acesso em: 30 abril de 2009.

CALLON, M.; COURTIAL, J.-P.; PENAN, H. **Cienciometría: la medición de la actividad científica: de la bibliometría a la vigilancia tecnológica**. Astúrias: Ediciones Trea, 1995.

GUIMARÃES, J. A. C.; LIBERATORE, G. Panorama del análisis documental de contenido en el cono sur americano. In: CARIDAD SEBASTIAN, M.; NOGALES FLORES, T. (Org.). **La información en la posmodernidad: la sociedad del conocimiento en España e Iberoamerica**. Madrid: Centro de Estudios Ramón Arce, 2004. p. 129-141.

HJÓRLAND, B. Domain analysis in information science:eleven approaches-traditional as well as innovative. **Journal of Documentation**, v. 58, n. 4, p. 422-462, 2002a.

\_\_\_\_\_. Epistemology and the socio-cognitive perspective um Information Science. **Journal of the American Society for Information Science and Technology**, v. 53, n. 4, p. 257-270, 2002b.

KERR, E. S. **Ketib**: um processo de representação de informações para textos complexos. 2003. Dissertação (Mestrado Profissional em Ciência da Computação) - Instituto de Computação da Universidade Estadual de Campinas, Campinas, 2003. Disponível em: <<http://libdigi.unicamp.br/document/?code=vtls000311557>>. Acesso em: 2 fev. 2006.

MACIAS-CHAPULA, C. A. O papel da infometria e da cienciometria e sua perspectiva nacional e internacional. **Ciência da Informação**, v. 27, n. 2, 1998. Disponível em: <http://www.ibict.br/cionline/>. Acesso em: 4 jan.2009.

MAI, J.-E. Analysis in indexing: document and domain centered approaches. **Information Processing and Management**, v. 41. p. 599-611, 2005.

NARIN, F.; OLIVASTRO, D.; STEVENS, K. S. Bibliometric theory, practice and problem. **Evaluation Review**, v. 18, n. 1, p. 65-76, 1994.

OKUBO, Y. **Bibliometric indicators and analysis of research systems**: methods and examples. Paris: OECD, 1997.

OLMEDA GÓMEZ, C.; PERIANEZ- RODRIGUEZ, A.; OVALLE-PERANDONES, M. A.; Estructura de las redes de colaboración científica entre las universidades españolas. **Ibersid 2008: revista de sistemas de información e comunicación**, p.129-140, 2008.

SPINAK, E. Indicadores cienciométricos. **Ciência da Informação**, Brasília, v. 27, n. 2, p. 141-148, 1998.

WASSERMAN, S.; FAUST, K. **Social network analysis**: methods and applications. Cambridge: Cambridge University Press, 1994.