

CO-CLASSIFICAÇÃO ENTRE ARTIGOS E PATENTES: um estudo da interação entre C&T na Biotecnologia Brasileira

Ana Maria Mielniczuk de Moura*
Sonia Elisa Caregnato**

RESUMO

Analisa a interação entre a ciência e a tecnologia (C&T) a partir de um enfoque cientométrico, utilizando a técnica de co-classificação, utilizada em outros estudos. Objetiva revelar o trânsito de pesquisadores e inventores da área de Biotecnologia no Brasil, no período de 2001 a 2005, nas esferas científica e tecnológica, a partir da correlação dos assuntos dos artigos e das patentes de sua autoria. O corpus constitui-se em 194 patentes e 2.584 artigos coletados na Base de Pedidos de Patentes do INPI e na *WebofScience*, respectivamente. A correlação entre os assuntos dos artigos e das patentes foi realizada por um especialista da área de Biotecnologia, a partir do mapeamento das categorias proposto por Glänzel e Schubert (2003) e dos códigos da Classificação Internacional de Patentes (CIP). Os resultados indicam que a produção intelectual na área da Biotecnologia no Brasil caracteriza-se por apresentar interação entre C&T, conforme verificado através da análise de co-classificação entre os dois tipos de documentos, atingindo um elevado percentual de interação.

Palavras-chave: Cientometria. Co-classificação. Artigo científico. Patente. Interação C&T.

* Doutora em Comunicação e Informação pelo PPGCOM/UFRGS. Professora Assistente do Departamento de Ciências da Informação/FABICO/UFRGS.
E-mail: ana.moura@ufrgs.br

** Doutora em Ciência da Informação pela Sheffield University, Inglaterra. Professora do Departamento de Ciências da Informação e do PPGCOM/UFRGS. Vice-coordenadora do PPGCOM/UFRGS.
E-mail: sonia.caregnato@ufrgs.br

I INTRODUÇÃO

Considera-se que artigos e patentes, ambos entendidos como suporte da informação, podem ser analisados a partir das mesmas técnicas cientométricas, pois possuem características análogas do ponto de vista do documento (autor/inventor, instituição/depositante, classificação bibliométrica/classificação oficial, etc), embora sob a perspectiva de produção, apresentam diferenças significativas (difusão/sigilo, troca/comercialização, etc.)

Os questionamentos que nortearam esse trabalho podem ser assim relacionados: quais as relações entre a produção científica dos pesquisadores da área da Biotecnologia e a sua

produção tecnológica? Os pesquisadores que registram patentes também publicam artigos? Os assuntos dos artigos se repetem nas patentes e vice e versa? Assim, procurou-se verificar se a produção científica, aqui caracterizada pelo artigo de periódico publicado pelos pesquisadores, apresenta relação com as patentes depositadas pelos mesmos autores, através da co-ocorrência dos assuntos constantes nas patentes e nos artigos, utilizando-se métodos cientométricos de análise de co-classificação. Considera-se que através destes aspectos possa-se inferir uma relação entre ciência e tecnologia.

A interação entre assuntos de patentes e publicações científicas, a partir da co-classificação, técnica cientométrica que possibilita a correlação entre diferentes classificações de assuntos, já foi

alvo de vários estudos (BASSECOULARD; ZITT, 2004; GLÄNZEL; MEYER, 2006a; LEYDESDORFF, 2008a; LEYDESDORFF, 2008c; TURNER et al., 1991). No entanto, ela recebe críticas por alguns desses autores. Entre as principais delas, destaca-se Bassecoulard e Zitt (2004), que se referem às diferenças entre os dois esquemas de classificação da ciência e tecnologia.

O grande volume de pesquisas e desenvolvimentos gerados principalmente no âmbito da universidade, no Brasil, resulta em uma expressiva produção intelectual de seus pesquisadores, formalizada, em maior grau, como artigos de periódicos, mas também como patentes. Essa produção, pela sua importância econômica, social e estratégica deve ser estudada através de várias óticas. Uma delas é aquela que se detém na interação entre Ciência e Tecnologia, por meio da colaboração entre pesquisadores e outra por meio da análise da co-ocorrência dos assuntos entre artigos e patentes.

A Biotecnologia foi selecionada para o estudo por apresentar uma intrínseca relação tanto com a Ciência como com a Tecnologia. Sua área de atuação, de maneira geral, envolve temas como transgênicos, as modificações das características de alimentos e de plantas pelo chamado melhoramento genético, a alteração artificial de características de plantas e de animais pela Biotecnologia aplicada, além da ampla aplicação na área da Saúde, através da criação de novas drogas e vacinas. (BRISOLLA et al., 1997). Ou seja, a biotecnologia utiliza-se dos conhecimentos científicos aliados à técnica necessária para intervenção e modificação na natureza, demonstrando ser uma área que se situa entre essas duas esferas, possibilitando que o estudo da interação entre a produção científica e tecnológica atinja os objetivos propostos.

Este artigo está organizado da seguinte forma: primeiramente, na seção dois, será abordado a cientometria e a técnica de co-classificação, utilizada neste estudo. Na seção três descreve-se a metodologia, seguida das conclusões do estudo, na seção quatro, e das referências utilizadas no trabalho, ao final.

2 CIENTOMETRIA E CO-CLASSIFICAÇÃO

Callon, Courtial e Penan (1993) definem o termo cientometria como os estudos direcionados à análise quantitativa da atividade

de pesquisadores e tecnólogos. A cientometria estuda os recursos e os resultados referentes à organização da produção do conhecimento e seu *savoir-faire*. Portanto, os autores mencionam que, recentemente, ela está exclusivamente interessada na análise dos documentos redigidos pelos pesquisadores e tecnólogos, ou seja, os artigos e patentes.

Algumas bases de dados armazenam ambas as formas de documentos, o que facilita ainda mais os estudos cientométricos. Autores como Narin (1994), Pavitt (1985), Bassecoulard e Zitt (2004) concordam com a aplicação destas técnicas em artigos e patentes.

Bassecoulard e Zitt (2004) discutem que a ligação entre ciência e tecnologia pode ser investigada por uma variedade de ferramentas baseadas na estrutura informétrica das publicações e patentes: citações, co-atividade, relações de classificação, abordagens léxicas ou abordagens híbridas. Porém, questionam a categoria que relaciona esquemas de classificação da ciência e tecnologia, dizendo que são basicamente incomensuráveis: classificações de publicações e patentes não são conectadas uma com a outra, pois as classificações de patentes são fortemente codificadas pelos escritórios de patentes, através de duas nomenclaturas oficiais: USPTO e International Patent Classification (IPC), não utilizadas pela nomenclatura científica, principalmente pelas bases de dados mais usadas em estudos bibliométricos: ISI e INIST Pascal.

É verdadeira a dificuldade de relacionar dois sistemas de classificação tão diferentes, e estudos cientométricos que queiram comparar artigos e patentes, no nível dos assuntos, serão beneficiados com a utilização de bases de dados que já englobem ambos os documentos, com a mesma classificação atribuída. No entanto, poucas são as bases de dados temáticas desse tipo. A base CA Chemical Abstracts é um exemplo: ela divide a Química em 80 seções, 80 códigos diferentes, que são atribuídos aos artigos e patentes que se relacionam com aquela seção.

Na conclusão dos autores que utilizaram a técnica de co-classificação (TODOROV¹, 1992; TODOROV, WINTERHAGEN², 1991, apud ROSTAING, 1996), aplicada à base de dados

1 TODOROV, R. Displaying Content of Scientific Journal: a co-heading analysis. *Scientometrics*, v.23, n.2, p.317-334, 1992.

2 TODOROV, R.; WINTERHAGEN, M. An overview of Mike Moravcsik's publication activity in Physics. *Scientometrics*, v.20, n.1, p.163-172, 1991.

Inspec, os códigos são considerados como palavras-chave, mas em nível de agregação muito mais elevado. Os autores estimam que seu emprego oferece vantagens consideráveis, pois não depende da linguagem do autor do artigo, nem dos limites de cobertura do SCI, circunscrito aos artigos de periódicos mais citados, e por ser mais objetivo que as palavras empregadas pelo autor, que são geralmente baseadas no julgamento do indexador.

Leydesdorff (2008c), ao comentar as diferenças entre os sistemas de classificação de artigos e de patentes, argumenta que o *status* deste indicador é diferente do indicador de classificação da ciência, principalmente devido ao crescimento do sistema de classificação de patentes em número de dígitos, o que amplia o entendimento humano, mas torna o sistema complexo, o que prejudica a reflexão sobre as dinâmicas entre as classificações e a identificação das estruturas cognitivas.

Quanto aos esquemas de classificação, Turner, Buffet, Laville (1991)³ estudaram, em nível macro, a verbalização utilizada no código de classificação internacional de patentes. Leydesdorff (2008a) utilizou a co-classificação para delinear as áreas de Nanoscience e Nanotechnology em 12 periódicos da área e patentes, a partir da nova categoria utilizada pela WIPO para a nanotecnologia, a YO1N, para identificar as relações entre os países a partir das relações entre as citações destes documentos. Em outro estudo, Leydesdorff (2008c) destaca que as classificações de patentes podem ser comparadas com as categorias de assunto atribuídas pelo ISI aos periódicos, embora o autor tenha concluído que países não são boas unidades de análises, porque os portfólios de patentes são muito semelhantes em países avançados. Sugere que sejam utilizadas as palavras dos títulos como uma melhor abordagem para visualizar a organização intelectual de patentes.

Meyer (2005) explora a relação entre patenteamento em nanotecnologia e ciência na Suécia, através dos inventores que também possuem artigos, aplicando a análise de redes de co-autoria. O autor apresenta as redes de co-invenção formadas entre universidade e indústria, e redes de co-classificação (análise

de co-ocorrência de frequência das subclasses IPC, relacionadas a diferentes setores da nanotecnologia). Em suas conclusões, o autor mostra que as análises da ligação do patenteamento em nanotecnologia para a ciência acadêmica são a base do estabelecimento de *relações* entre universidades e indústria.

3 METODOLOGIA

A abordagem utilizada neste estudo é quantitativa a partir de uma perspectiva cientométrica. Entre os métodos, foram utilizados métodos estatísticos aliados a Análise de Redes Sociais – ARS, que deram subsídio para entender os relacionamentos entre entidades e sujeitos. As entidades analisadas neste estudo são os artigos de periódicos e patentes. Os atores são aqui identificados como autores (responsáveis intelectuais dos artigos), inventores (responsáveis intelectuais das patentes) ou co-ativos (pesquisadores que possuem artigos e patentes) quando aplicado àqueles que transitam pela Ciência e Tecnologia.

Os sujeitos deste estudo constituem-se dos pesquisadores brasileiros que possuem tanto pedidos de patentes como artigos publicados na área de Biotecnologia. A amostra foi formada por aqueles que possuem pelo menos um pedido de patente depositado na Base de Patentes do INPI e um artigo indexado na base *WebofScience*.

O *corpus* constitui-se pela produção intelectual desses sujeitos, formalizada por artigos de periódicos e patentes, identificados nas bases *WebofScience* e Base de Patentes do INPI, respectivamente. O período selecionado para a análise das publicações são os anos de 2001 a 2005. Os anos de 2006 e 2007 foram excluídos porque o período de sigilo de uma patente é de 18 meses e somente a partir deste prazo os seus dados são divulgados ao público. Assim, a coleta de dados, que aconteceu em 2008, não poderia contemplar os dois anos anteriores.

A produção tecnológica, neste estudo representada pelas patentes registradas pelos pesquisadores, foi identificada na Base de Patentes do INPI, através da busca pelos códigos de classificação CIP (Classificação Internacional de Patentes). (WORLD INTERNATIONAL PROPERTY ORGANIZATION, 2007).

Foram utilizados os códigos relacionados à área da Biotecnologia, conforme a Organisation for

³ TURNER, W.A.; BUFFET, P.; LAVILLE, F. LEXITRAN for an easier public access to patent databases. *World Patent Information*, v. 13, n. 2, p. 81–90, 1991. APUD BASSECOULARD; ZITT, 2004.

Economic Co-operation and Development (2005). (Science, Technology and Industry Scoreboard 2005 - Towards a Knowledge based economy): A01H, A61K, C02F, C07G, C07K, C12M, C12N, C12P, C12Q, C12S e G01N. Códigos adicionais foram recuperados quando acompanharam este grupo de códigos selecionados, mas não fizeram parte das análises.

O período de depósito das patentes foi de 01/01/2001 até 31/12/2005. A partir dos resultados encontrados, selecionaram-se os pedidos identificados como tendo pelo menos um inventor brasileiro. Ao final, contabilizaram-se 194 patentes nacionais e 2331 patentes estrangeiras, sendo que o corpus foi constituído dos primeiros.

A produção científica dos inventores brasileiros identificados como detentores de patentes na área da Biotecnologia, foi recuperada e identificada no SCI da *WebofScience*. A coleta iniciou-se logo após o término da coleta das patentes, em janeiro de 2008, e foi concluída no final de fevereiro de 2008. O total de artigos coletados foi de 4.288, e, após a supressão dos registros duplicados, este número foi reduzido para 2.584.

Padronizaram-se os assuntos conforme as categorias propostas por Glänzel e Schubert (2003), que organizaram os assuntos indexados no ISI em 17 categorias. Em segundo lugar, devido à complexidade da atividade de correlacionar dois tipos de classificação tão distintos, solicitou-se a participação de um especialista da área para relacionar as codificações da CIP com o esquema de classificação de assuntos utilizado. Solicitou-se a um especialista da área de Biotecnologia, pesquisador com bolsa de produtividade e detentor de pedido de patente no INPI que relacionasse as categorias de assunto dos artigos aos códigos CIP, identificadores dos assuntos das patentes.

Em terceiro lugar, após a correlação realizada pelo especialista, foi possível, através de uma programação, aplicar esta correlação no banco de dados, procedimento que possibilitou identificar em quais casos a co-classificação ocorria entre artigos e patentes. Além disso, foi possível identificar quais os casos em que tanto a co-autoria e co-invenção como a co-classificação ocorriam concomitantemente, dando indícios de que estes documentos artigo e patente eram

representantes de alta interação entre ciência e tecnologia. Foram aplicadas análises estatísticas e ARS para esta análise.

4 RESULTADOS

Analisando-se a distribuição dos assuntos dos periódicos nos artigos após a conversão para as categorias de Glänzel e Schubert (2003), observa-se que alguns assuntos sobressaem-se, como Biociências e Biologia. A distribuição dos assuntos em tabela (Tabela 1), com os percentuais que cada categoria apresenta, e gráfico (Figura1) possibilita a observação da participação de cada categoria.

Tabela 1 - Artigos por Assunto

Assunto	Artigos	%
Biociências	1136	44,0
Biologia	1017	39,4
Química	766	29,6
Pesquisa Biomédica	409	15,8
Med. Clin. Experimental I	319	12,3
Medicina Clin. Experimental II	277	10,7
Física	169	6,5
Agricultura e Meio Ambiente	165	6,4
Engenharia	45	1,7
Neurociências e Comportamento	39	1,5
Ciências Multidisciplinares	20	0,8
Matemática	20	0,8
Ciências Sociais I	8	0,3
Geociências e Ciência Espacial	3	0,1
Nanociência e Nanotecnologia	2	0,1
Nr. Total de Artigos	2.584	

Fonte: Dados do autor.

Nota: A soma dos itens na coluna artigos não é igual ao no total de artigos, pois alguns artigos têm mais do que um assunto.

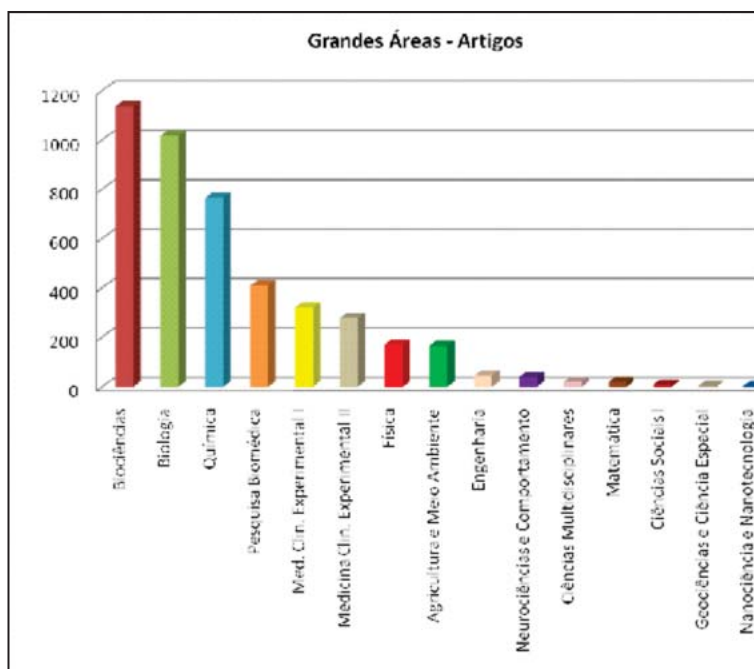


Figura 1 – Áreas Artigos

Fonte: Dados do autor.

Observa-se que apenas 15 das 17 categorias pontuaram, ficando de fora as categorias Artes e Geociências.

A organização dos assuntos nos artigos está descrita na Tabela 1, onde se pode observar o percentual de participação de cada assunto, assim como o seu percentual cumulativo.

Observa-se na Tabela 1 e na Figura 1 que os artigos publicados pelos sujeitos são predominantemente da área de Biociências, seguidos da Biologia e Química. Biociências, conforme a classificação de Glänzel e Schubert (2003), trata da Biologia geral, celular e subcelular e Genética, e em suas sub-classificações, abrange a Bioquímica, Biofísica e Biologia Molecular e Biologia Celular. Estas são áreas estreitamente relacionadas com a Biotecnologia.

Pesquisa Biomédica e Medicina Clínica e Experimental também apresentam destaque, o que significa que muitas das pesquisas nesta área são resultados de pesquisas aplicadas à bioengenharia, biomateriais e farmacologia, esta última responsável pelo desenvolvimento de medicamentos e fármacos, estreitamente relacionada ao patenteamento. As sub-áreas da Medicina Clínica dão conta de pesquisas em Endocrinologia e Metabolismo, Oncologia e Imunologia, área também bem envolvida com o

desenvolvimento de medicamentos e fármacos, além de processos de tratamento.

Analisando o percentual cumulativo, constata-se que 89,3% das publicações científicas dos co-ativos estão relacionadas às áreas ligadas a ciências da vida, pesquisa biomédica e, conseqüentemente, ao desenvolvimento de medicamentos e fármacos.

A Classificação Internacional de Patentes (CIP), atualmente em sua 8ª edição, apresenta as seguintes seções (INSTITUTO NACIONAL DE PROPRIEDADE INTELECTUAL, 2008): a) Necessidades Humanas; b) Operações de Processamento; Transporte; c) Química e Metalurgia; d) Têxteis e Papel; e) Construções Fixas; f) Engenharia Mecânica; Iluminação; Aquecimento; Armas; Explosão; g) Física; h) Eletricidade.

Cada uma dessas grandes seções possuem sub-divisões, e a indexação da patente em uma delas pode ser feita com os seguintes enfoques: de acordo com a descrição (estrutura, composição, etc.) da matéria descrita; e/ou de acordo com a aplicação/finalidade da matéria descrita.

No gráfico a seguir (Figura 2), apresenta-se a distribuição das patentes em estudo de acordo com a CIP:

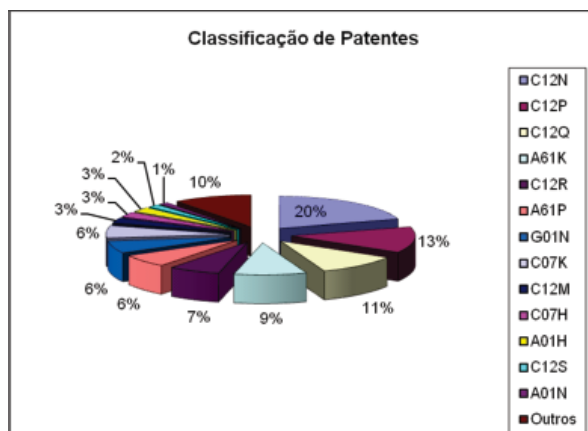


Figura 2 – Códigos de Classificação de Patentes – CIP

Fonte: Dados do autor.

Observa-se que as patentes estão classificadas mais frequentemente na Classe C, que se refere à Química e Metalurgia, seguida da classe A, Necessidades Humanas. 22% das patentes apresentam códigos com uma participação não significativa (< 3). Em destaque, os códigos C12N, C12P, C12Q e A61K. No Quadro 1, apresentam-se os códigos de classificação⁴ e o número de patentes encontradas na produção estudada, junto com a definição de cada código, conforme a CIP.

Código	Patentes	Definição
C12N	82	microorganismos ou enzimas; suas composições; propagação, preservação, ou manutenção de microorganismos ou tecido ; engenharia genética ou de mutações; meios de cultura
C12P	53	processos de fermentação ou processos que utilizem enzimas para sintetizar uma composição ou composto químico desejado ou para separar isômeros óticos de uma mistura racêmica.
C12Q	43	processos de medição ou ensaio envolvendo enzimas ou microorganismos. composições; controle responsivo a condições do meio nos processos microbiológicos ou enzimáticos
A61K	38	preparações para finalidades médicas, odontológicas ou higiênicas
C12R	27	relativo a microorganismos
A61P	26	atividade terapêutica de compostos químicos ou preparações medicinais
G01N	25	investigação ou análise dos materiais pela determinação de suas propriedades químicas ou físicas
C07K	25	Peptídeos
C12M	11	aparelhos para enzimologia ou microbiologia
C07H	10	açúcares; seus derivados; nucleosídeos; nucleotídeos; ácidos nucleicos
A01H	10	novas plantas ou processos para obtenção das mesmas; reprodução de plantas por meio de técnicas de cultura de tecidos
C12S	6	processos que utilizam enzimas ou microorganismos para liberar, separar ou purificar um composto ou uma composição pré-existente
A01N	5	conservação de corpos de seres humanos ou animais ou plantas ou partes dos mesmos; biocidas: desinfetantes, pesticidas, herbicidas
Outros	41	
Total	402	

Quadro 1 – Códigos das Patentes e sua Definição

Fonte: Dados do autor.

⁴ Um documento de patente pode conter uma ou mais classificações.

Neste estudo, o código C12N predomina, com 20% de participação. Fortes e Lage (2006), ao analisarem as patentes da área da Biotecnologia, selecionaram o código C12N como representante da área, por representar bem a engenharia genética. Eles identificaram, no triênio 1998 a 2000, um predomínio das patentes neste código e suas subdivisões, o que mostra o rápido crescimento do número de depósitos em engenharia genética e tecnologia de DNA recombinante, mostrando um reflexo

da corrida genômica que ocorre hoje em escala mundial. Observa-se que os códigos predominantes neste estudo referem-se à engenharia genética, processos de fermentação, preparações medicinais (caso dos medicamentos) e microorganismos. Em outros, encontram-se 41 patentes que foram classificadas em códigos que tiveram uma ocorrência não significativa.

A Figura 3 apresenta a rede formada pelos relacionamentos entre os assuntos das patentes, a partir dos códigos atribuídos a elas.

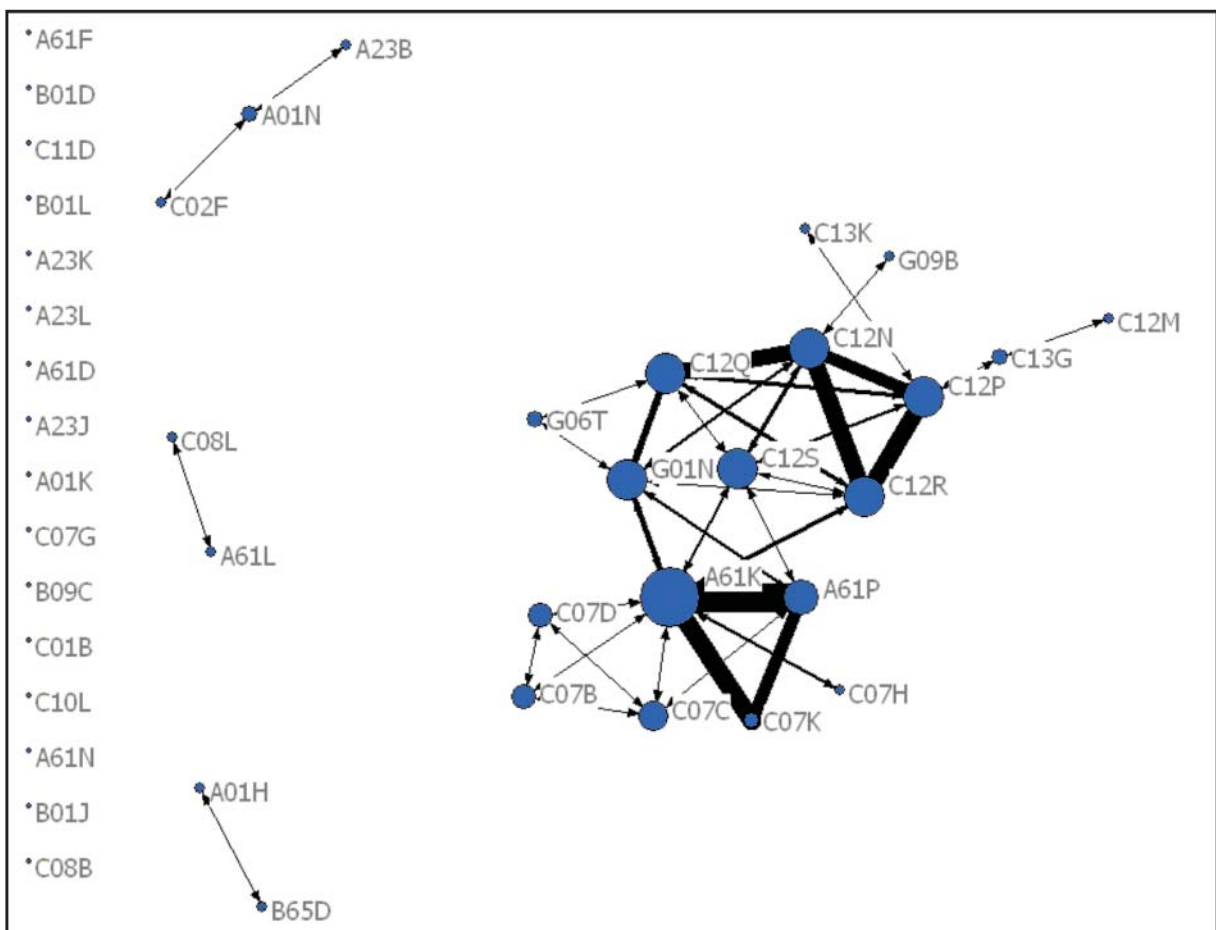


Figura 3 - Grafo Rede Geral Códigos de Assuntos das Patentes
Fonte: Dados do autor.

Observa-se a formação de dois *clusters* definidos, que estabelecem uma interligação entre si. O primeiro, com laços intensos entre os códigos C12N, C12Q, C12P, C12R, sendo os três primeiros, relacionados a microrganismos, com o enfoque da Engenharia Genética e meios de Cultivo ou processos de medição ou análise, e C12P, aos processos de fermentação ou processos que utilizam enzimas para a síntese de um composto químico ou composição química, ou para a separação de isômeros ópticos. O segundo *cluster*, formado pelos laços também intensos entre A61K, A61P, C07K e C07H, códigos relacionados à preparações medicinais, medicamentos, incluindo o envolvimento de peptídeos e ácidos nucleicos. Identifica-se a interligação entre estes dois *clusters* a partir dos nós G01N e C12R.

A predominância dos códigos justifica-se, pois a coleta das patentes foi feita a partir dos códigos relacionados à área da Biotecnologia, definidos pela Organisation for Economic Co-operation and Development (2005). O que se pode concluir é que os

códigos que apresentam laços fortes entre si referem-se a abordagens distintas do assunto das patentes, pois a indexação da patente pode ser de acordo com a descrição da matéria, a ação que ela exerce, sua função ou com a sua aplicação.

À esquerda do grafo, encontram-se os códigos isolados, pois foram recuperados junto aos códigos selecionados para a coleta, mas não fazem parte da relação indicada pela OECD para a área da Biotecnologia.

As redes diádicas e triádicas, também localizadas à esquerda do grafo, indicam que esses códigos não se interligam com os códigos com maior centralidade na rede, também pelo fato da maioria deles não fazerem parte da lista da OECD.

O resultado da análise do especialista sobre a correlação dos códigos dos assuntos das patentes com os assuntos dos artigos (Tabela 2) ordena na primeira coluna o código referente à área do assunto no ISI e na segunda, os códigos de classificação das patentes correspondentes:

Tabela 2– Relação entre Assuntos dos Artigos com os Códigos das Patentes

Código assunto ARTIGOS	Código assunto PATENTES
Agricultura e Meio Ambiente	A01H
Biologia	C12N,C12S
Biociências	A01H,C07K,C12N,C12P,C12Q,C12S
Pesquisa Biomédica	A61K
Med. Clin. Experimental I	NÃO SE APLICA
Med. Clin. Experimental II	NÃO SE APLICA
Neurociências e Comportamento	NÃO SE APLICA
Química	C02F,C07G,G01N
Física	NÃO SE APLICA
Geociência e Ciência Espacial	NÃO SE APLICA
Engenharia	C12M
Matemática	NÃO SE APLICA
Ciências Sociais I	NÃO SE APLICA
Ciências Sociais II	NÃO SE APLICA
Artes & Humanidades	NÃO SE APLICA
Ciências Multidisciplinares	A01H,A61K,C12N

Fonte: Dados do autor.

Observa-se que o assunto Biociências (Categoria 3) apresentou um maior número de códigos relacionados. Comparando com os resultados da análise dos assuntos dos artigos, este obteve o maior percentual, significando que o assunto predominante dos artigos publicados pelos pesquisadores da área de Biotecnologia está relacionado a Biociências. Outro assunto que apresentou vários códigos relacionados nas patentes é a Química, também predominante nos artigos. Isto indica uma convergência entre os assuntos dos artigos e das patentes, a

ser confirmado nas análises a seguir. Entre os assuntos que não se aplicam aos códigos das patentes estão aqueles mais ligados às áreas das ciências sociais, artes e medicina clínica.

Para visualizar esta interação entre artigos e patentes a partir dos assuntos compartilhados, criou-se uma matriz dos vínculos entre estes documentos. A Figura 4 apresenta a rede formada pelas patentes e artigos relacionados ao seu assunto, elaborada a partir da matriz, e visualizada através do software NETDRAW.

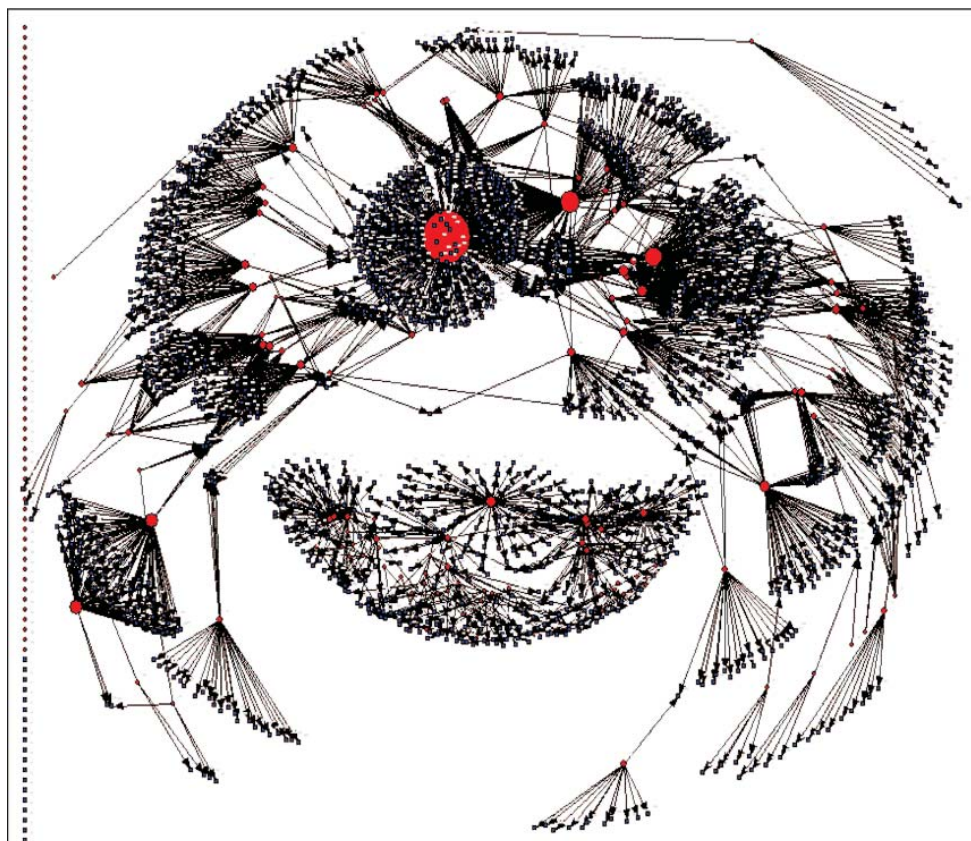


Figura 4 – Grafo Rede geral patentes e artigos que possuem mesmo assunto

Fonte: Dados do autor

Observam-se nesta rede as patentes e os artigos que tratam do mesmo assunto, em duas configurações: uma rede maior, interligada, e uma menor, no centro, desconectada da maior. Apresenta 2.438 laços, e os nós maiores, em vermelho, são aqueles que possuem maior grau de centralidade. É o caso da patente 68, que apresenta 361 artigos do mesmo assunto da patente. O próximo nó que se destaca é a patente 76, com 129 artigos do mesmo assunto, que apresenta 144

artigos relacionados pela co-autoria. Outro caso a ser destacado é a patente 45, com 103 artigos do mesmo assunto, que apresenta 129 artigos relacionados pela co-autoria. A patente 138, com 70 artigos relacionados ao seu assunto, apresenta 73 artigos relacionados pela co-autoria. A patente 159 também se destaca, pois apresenta 66 artigos com o mesmo assunto. Comparando com a análise de co-autoria, esta patente apresenta 70 artigos relacionados pela co-autoria.

Em relação aos assuntos que mais possuem laços, destacam-se os que estão vinculados às patentes citadas acima, que apresentam maior grau de centralidade: C12N, A61K, G01N, C07K, C12P, C07H, C12Q. Comparando com a análise de assuntos das patentes e dos artigos realizadas anteriormente, observa-se que estes códigos figuram entre os primeiros lugares na classificação de patentes (Figura 2). E na classificação dos assuntos dos artigos, as áreas que mais apareceram foram, nesta ordem: Biotecnologias, Biologia, Química e Pesquisa Biomédica (Tabela 2).

A Tabela 3 apresenta as patentes com percentual de $\geq 16\%$ de co-ocorrência entre co-classificação e co-autoria, sendo que os que possuem maior percentual aparecem no início da tabela. Ela está organizada nas seguintes colunas: Código da

Patente: número da patente atribuído no banco de dados da pesquisa; Co-ativos: número de co-ativos inventores da patente; Total Artigos: número total dos artigos publicados pelos co-ativos; Nr. Art. em co-autoria: número de artigos que tiveram pelo menos dois co-ativos; Co-autoria total: número de artigos em que a co-ocorrência dos co-ativos em artigos e patente é total; % Co-autoria total: o percentual da correlação total na co-autoria; Nr. Art. mesmo assunto patente: a quantidade de artigos que apresentam o mesmo assunto da patente; % Co-classif.: entre o total de artigos, qual o percentual de co-classificação; Co-classif. e co-autoria: número de artigos que apresentam correlação entre a co-classificação e co-autoria; % Co-classif. e co-autoria: percentual de co-classificação e co-autoria dos artigos identificados por já apresentarem correlação na co-autoria.

Tabela 3 – Artigos com $\geq 16\%$ Co-ocorrência entre Co-classificação e Co-autoria

Código Patente	Co-ativos	Artigos	Em co-autoria	Co-autoria total	% Co-autoria total	Nr. art. mesmo assunto patente	% Co-classif. (do total de artigos)	Co-classif. e co-autoria	% Co-classif. e co-autoria
176	2	3	3	3	100	3	100	3	100
119	4	3	1	1	33,3	1	33,3	1	100
141	2	7	5	5	71,4	2	28,6	2	100
4	3	7	3	3	42,9	1	14,3	1	100
180	3	18	3	1	5,6	1	5,6	1	100
47	2	7	5	5	71,4	5	71,4	4	80
28	2	7	5	5	71,4	7	100	5	71,4
184	2	3	2	2	66,7	3	100	2	66,7
127	2	10	5	5	50	9	90	5	55,6
116	2	7	4	4	57,1	6	85,7	3	50
64	2	12	6	6	50	10	83,3	5	50
65	2	12	6	6	50	6	50	3	50
92	2	5	3	3	60	2	40	1	50
102	2	11	5	5	45,5	11	100	5	45,5
85	2	23	6	6	26,1	12	52,2	5	41,7
27	2	10	2	2	20	5	50	2	40
41	2	25	6	6	24	25	100	6	24
115	2	18	6	6	33,3	17	94,4	6	35,3
86	3	21	16	7	33,3	21	100	7	33,3
5	2	11	2	2	18,2	6	54,5	2	33,3
118	2	25	5	5	20	16	64	4	25
20	2	50	10	10	20	20	40	4	20
121	3	16	11	5	31,3	11	68,8	2	18,2
98	2	22	3	3	13,6	12	54,5	2	16,7
174	2	11	4	4	36,4	6	54,5	1	16,7
145	4	61	11	5	8,2	31	50,8	5	16,1

Fonte: Dados do autor.

A partir da análise dos dados, observou-se que 131 patentes apresentam artigos de mesmos autores relacionados ao seu assunto. Entre esses artigos, cinco possuem 100% de correlação de autoria e de assunto com as patentes (Tabela 3).

Para uma melhor interpretação desses dados, deve-se observar que, na coluna intitulada % Co-classif., tem-se o percentual de co-classificação calculado a partir do total de artigos que estão vinculados à patente (terceira coluna, Total de Artigos), a partir da co-autoria. Assim, pode-se observar que a maioria das patentes nesta tabela apresenta um percentual maior de correlação nos assuntos do que nas co-autorias. A patente 47 apresenta convergência entre o percentual de co-autoria e co-classificação, pois cinco dos sete artigos são da mesma autoria e quatro destes cinco apresentam a mesma co-autoria e co-classificação. A patente 28 também apresenta convergência na co-autoria, num

percentual de 71,4 de correlação. Na correlação dos assuntos, este percentual sobe para 100%, significando que todos os sete artigos publicados pelos co-ativos apresentam similaridade com o assunto da patente destes co-ativos. Assim, a correlação total entre co-autoria e co-classificação desta patente atingiu 71,4. Outros casos desta natureza apareceram, mostrando que a correlação nos assuntos é mais intensa do que na co-autoria, talvez pelo fato dos assuntos terem sido organizados em grandes áreas. Destaca-se que um mesmo código de classificação foi atribuído pelo especialista a mais de uma categoria de assunto.

A Tabela 4 mostra quais patentes possuem maior convergência entre co-autoria e co-classificação, convergência sendo aqui entendida como uma maior semelhança entre os percentuais atingidos nestas categorias.

Tabela 4 – Patentes com Artigos de mesma Co-classificação e Co-autoria

Patente	Co-ativos	Nr. Art.	% mesma co-autoria	% mesma co-classificação
176	2	3	100	100
141	2	7	71,4	28,6
4	3	7	42,9	14,3
119	4	3	33,3	33,3
180	3	18	5,6	5,6
47	2	7	71,4	71,4
28	2	7	71,4	100
184	2	3	66,7	100
127	2	10	50	90
92	2	5	60	40
116	2	7	57,1	85,7
64	2	12	50	83,3
65	2	12	50	50
102	2	11	45,5	100
85	2	23	26,1	52,2
41	2	25	40	100
27	2	10	20	50
115	2	18	33,3	94,4
86	3	21	33,3	100
5	2	11	18,2	54,5
118	2	25	20	64
20	2	50	20	40
121	3	16	31,3	68,8
174	2	11	36,4	54,5
98	2	22	13,6	54,5
145	4	61	8,2	50,8
66	3	8	12,5	87,5
37	2	82	2,4	8,5
87	2	35	11,4	94,3
126	3	15	6,7	60
136	4	22	9,1	90,9

Fonte: Dados do autor.

Constata-se a predominância dos assuntos das patentes nos artigos, independente da total co-ocorrência de co-autoria: as patentes 28 e 184 apresentam 100% de co-classificação, embora em co-autoria, alcancem 71,4% e 66,7%, respectivamente. Outro destaque neste sentido é a patente 102, que apresenta 100% de co-classificação e apenas 45,5% de co-autoria. Isto significa que a interação acontece tanto em co-autorias como em assuntos. Mas os temas de pesquisa estudados pelos co-ativos tanto em patentes como em artigos possuem similaridade. Isto demonstra que os autores seguem uma linha de pesquisa, mostrando a especialização na área da Biotecnologia, apesar do seu caráter interdisciplinar. Cabe lembrar que as categorias de assuntos dos artigos são gerais, facilitando, portanto, o encaixe nas grandes classes.

Entre as 157 patentes que possuem co-ativos vinculados, 131 delas relacionam-se a artigos que possuem o mesmo assunto, alcançando um percentual de 83,4% de interação entre Ciência e Tecnologia a partir da análise da co-classificação realizada. Ao comparar o total de patentes em todo o estudo (194), incluindo as que não possuem co-ativos vinculados, este percentual diminui um pouco, resultando em 67,5% de interação.

Como demonstram as análises, conclui-se que existe interação entre Ciência e Tecnologia na área da Biotecnologia. A obtenção destes resultados só foi possível a partir da aplicação das técnicas de Cientometria, que envolveram as análises de co-classificação entre os assuntos dos artigos e das patentes.

5 CONCLUSÕES

Entre as 157 patentes que possuem co-ativos vinculados, 131 delas relacionam-se a artigos que possuem o mesmo assunto, alcançando um percentual de 83,4% de interação entre Ciência e Tecnologia a partir da análise da co-classificação realizada. Quando é comparado o total de patentes em todo o estudo (194), incluindo as que não possuem co-ativos vinculados, este percentual diminui para um índice de 67,5% de interação.

Observa-se que as patentes que mais possuem convergência entre co-autoria e co-classificação são aquelas que possuem menor

número de co-ativos. Outra constatação diz respeito à ocorrência dos assuntos das patentes nos artigos, mesmo quando não há repetição integral da co-autoria. Isto significa que a interação acontece tanto em co-autorias como em assuntos, e as parcerias formadas por ocasião da elaboração da patente se repetem em parte, como já constatado na análise de co-autoria. Mas os temas de pesquisa estudados pelos co-ativos tanto em patentes como em artigos possuem similaridade, o que demonstra que os autores seguem uma linha de pesquisa.

Conclui-se que a co-classificação é uma técnica difícil de aplicar em tipos de documentos diferentes e que apresentam particularidades na indexação dos seus assuntos. As categorias de assuntos dos artigos são gerais, e em menor número, já os códigos da Classificação Internacional de Patentes são específicos e totalizam cerca de 70 mil. Além disso, a dificuldade na correlação entre os dois tipos de assuntos exige a participação de um especialista da área. Assim, conclui-se que a análise de co-classificação pode ser utilizada, mas com o apoio de um especialista da área a ser estudada e, preferencialmente, que seja utilizada uma base de dados que indexe os dois tipos de documentos, utilizando as mesmas categorias de assunto.

Considera-se que o estudo da interação entre ciência e tecnologia na área da Biotecnologia no Brasil mostrou-se efetivo a partir das análises cientométricas de co-classificação aplicadas aos documentos de artigos e patentes, mostrando que a cientometria é uma metodologia adequada para este tipo de pesquisa, assim como as técnicas empregadas.

Considera-se de fundamental importância a existência de bases que arrolem tanto a produção científica como tecnológica, para que estudos desta natureza sejam mais factíveis. A agilidade no processo de realização de pesquisas desta natureza mostra-se necessária, pois facilitará o processo decisório de instituições de fomento e subsidiará as políticas públicas para a ciência e tecnologia no Brasil. Além disso, permitirá aos pesquisadores a recuperação e utilização dos dois tipos de documentos nas suas pesquisas.

CO-CLASSIFICATION BETWEEN ARTICLES AND PATENTS:

a study of the interaction between S & T on Brazilian Biotech

Abstract

The article examines the interaction between science and technology (S&T) from a scientometrics perspective, using the technique of co-classification, also used in other studies. It aims at showing the transit of researchers and inventors from the area of Biotechnology in Brazil, from 2001 to 2005, through the scientific and technological spheres by means of the correlation between articles and patents of their own. The corpus is made up of 194 patents and 2584 articles, collected from the INPI Applications for Patents database and the Web of Science, respectively. The correlation between the subjects of the articles and patents was done by a specialist in the area of biotechnology, by mapping the categories proposed by Glänzel and Schubert (2003) and the codes of the International Patent Classification (IPC). The results indicate the area of Biotechnology in Brazil is characterized by an interaction between S&T, as shown by the co-classification between the two types of documents produced by the researchers, reaching a high percentage of interaction.

Key-words:

Scientometrics. Co-classification. Scientific article. Patent. Interaction S&T.

Artigo recebido em 25/09/2009 e aceito para publicação em 11/07/2010

REFERÊNCIAS

BASSECOULARD, Elise; ZITT, Michel. Patents and Publications: the lexical connection. In: MOED, H. F.; GLÄNZEL, W.; SCHMOCH, U. (Eds.) **Handbook of Quantitative Science and Technology Research**. Netherlands: Kluwer Academic Publishers, 2004. p. 665-694.

BRISOLLA, S. et al. As Relações Universidade-empresa-governo: um estudo sobre a Universidade Estadual de Campinas (Unicamp). **Educação & Sociedade**, Brasília, v. 18, n.61, dez. 1997.

CALLON, Michel; COURTIAL, Jean-Pierre; PENAN, Hervé. **La Scientométrie**. Paris: Presses Universitaires de France, 1993.

FORTES, Maria Hercília Paim; LAGE, Celso Luiz Salgueiro. Depósitos Nacionais de Patentes em Biotecnologia, subclasse C12N, no Brasil de 1998 a 2000. **Biotemas**, Florianópolis, v.19, n. 1, p. 7-12, mar. 2006.

GLÄNZEL, Wolfgang; MEYER, Martin. Patents cited in the scientific literature: An exploratory study of 'reverse' citation relations. **Scientometrics**, Dordrecht, v.58, n. 2, p. 415-28, 2003.

GLÄNZEL, Wolfgang; SCHUBERT, András. A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. **Scientometrics**, Dordrecht, v. 56, n. 3, 2003, p. 357-367.

LEYDESDORFF, Loet. The delineation of nanoscience and nanotechnology in terms of journals and patents: A most recent update. **Scientometrics**, Dordrecht, v. 76, n. 1, jul. 2008a.

LEYDESDORFF, Loet. Patent classifications as indicators of intellectual organization. **Journal of the American Society for Information Science and Technology**, v. 59, n.10, p.1582-1597, Aug. 2008c.

MEYER, Martin. **Nanotechnology in Sweden: tracking patenting activity & links between nanotech firms and swedish science**. Helsinki: **Knowledge Flows**, 2005. Report to the Royal Swedish Academy of Engineering Sciences, 2005. Disponível em: <<http://www.iva.se/upload/Verksamhet/Projekt/Nano/Nanotechnology%20in%20Sweden%20060119.pdf>> Acesso em: 06 nov. 2007.

NARIN, F. Patent bibliometrics. **Scientometrics**, Dordrecht, v.30, n.1, p.147-155, 1994.

PAVITT, K. Patent statistics as indicators of innovative activities : possibilities and problems. **Scientometrics**, Dordrecht, v.7, p.77-99, 1985.

ROSTAING, Hervé. La Bibliométrie et ses techniques. Toulouse/Marseille: Édition Sciences de la Societé/Centre de Recherche Rétrospective de Marseille, 1996. (**Sciences de la Societé, Suplemento** n. 38).

STOKES, Donald E. O Quadrante de Pasteur: a ciencia básica e a inovação tecnológica. Campinas: Editora da Unicamp, 2005.

WORLD INTERNATIONAL PROPERTY ORGANIZATION. International Patent Classification. Versão 2007.01. Disponível em: <<http://www.wipo.int/classifications/ipc/ipc8trans/es/ipcpub/>>. Acesso em: 09 jul. 2009.