

ISSN 0103-3786
VOLUME 26 NÚMERO 3
SETEMBRO/DEZEMBRO 2014

Trans**Informação**

FUNDADA EM 1989

Editor / Editor

Prof. Dr. Rogério Eduardo Rodrigues Bazi (PUC-Campinas, SP, Brasil)

Editora Associada / Associate Editor

Prof. Dra. Nair Yumiko Kobashi (USP, São Paulo, SP, Brasil)

Editora Gerente / Manager Editor

Maria Cristina Matoso (PUC-Campinas, SP, Brasil)

Conselho Editorial / Editorial Board

Prof. Dr. Aldo de A. Barreto (IBICT - UFRJ, Rio de Janeiro, RJ, Brasil)

Prof. Dra. Henriette Ferreira Gomes (UFBA, Salvador, BA, Brasil)

Prof. Dra. Isa Maria Freire (UFPB, João Pessoa, PB, Brasil)

Prof. Dra. Lígia Café (UFSC, Florianópolis, SC, Brasil)

Prof. Dr. Luís Fernando Sayão (CNEN, Rio de Janeiro, RJ, Brasil)

Prof. Dra. Maria de Fátima G. M. Tálamo (USP, São Paulo, SP, Brasil)

Prof. Dra. Maria Inês Tomaél (UEL, Londrina, PR, Brasil)

Prof. Dra. Maria Nélide González de Gómez (IBICT, Brasília, DF, Brasil)

Prof. Dr. Raimundo Nonato Macedo dos Santos (UFPE, Recife, PE, Brasil)

Prof. Dra. Rosali Fernandez de Souza (IBICT, Brasília, DF, Brasil)

Conselho Editorial Internacional / International Editorial Board

Prof. Dr. Antonio García Gutiérrez (Universidade de Sevilla, Sevilla, Espanha)

Prof. Dra. Fernanda Ribeiro (Universidade do Porto, Porto, Portugal)

Prof. Dr. Juan Carlos Fernández Molina (Universidade de Granada, Granada, Espanha)

Prof. Dr. Pierre Fayard (Universidade de Poitiers, Poitiers, França)

Prof. Dr. Yves-François Le Coadic (Conservatoire Nat. des Arts et Metiers, Bordeaux, França)

Normalização / Normalization

Bibliotecárias / Librarians

Andréa Ribeiro Alves Bonfim (PUC-Campinas, SP, Brasil)

Maria Cristina Matoso (PUC-Campinas, SP, Brasil)

Apoio Administrativo / Administrative Support

André Gustavo Tomaz dos Santos (PUC-Campinas, SP, Brasil)

INDEXAÇÃO / INDEXING

SciELO, Web of Science, JCR Social Science, Scopus, Latindex, Clase FI=0,083
Qualis A1

Copyright © Transinformação

É permitida a reprodução parcial desde que citada a fonte. A reprodução total depende da autorização da Revista.

Partial reproduction is permitted if the source is cited. Total reproduction depends on the authorization of the Transinformação.

O Conselho Editorial não se responsabiliza por conceitos emitidos em artigos assinados.

The Board of Editors does not assume responsibility for concepts emitted in signed articles.

Transinformação é uma revista especializada, com periodicidade quadrimestral, aberta a contribuições da comunidade científica nacional e internacional, editada pela Faculdade de Biblioteconomia, Centro de Ciências Humanas e Sociais Aplicadas, Pontifícia Universidade Católica de Campinas. Fundada em 1989, publica artigos que contribuem para o estudo e o desenvolvimento científico da Ciência da Informação, da Biblioteconomia, da Arquivologia, da Museologia e de áreas afins.

Transinformação is a specialized journal, issued three times a year, and open to contributions from the domestic and international scientific communities. It is edited by the School of Library Science, Center of Applied Human and Social Sciences, Pontifical Catholic University of Campinas. Founded in 1989, it publishes articles that contribute to the study and scientific development of Information Science, Library Science, Archival Science, Museology and related areas.

COLABORAÇÕES / CONTRIBUTIONS

Os manuscritos devem ser enviados à Secretaria da Revista via Portal de Periódico, <<http://www.puc-campinas.edu.br/periodicocientifico>> conforme as "Instruções aos Autores", publicadas no final de cada fascículo. *All manuscripts should be sent to the Transinformação' Office <<http://www.puc-campinas.edu.br/periodicocientifico>> and should comply with the "Guide for Authors", published in the end of each issue.*

CORRESPONDÊNCIA / CORRESPONDENCE

Toda a correspondência deve ser enviada à Transinformação no endereço abaixo:

All correspondence should be sent to Transinformação at the address below:

Núcleo de Editoração SBI

Prédio da Antiga Reitoria Sala 8 - Campus I

Rod. Dom Pedro I, km 136 - Pq. das Universidades

13086-900 - Campinas - SP

Fone/Fax: 55 (19) 3343-7401

E-mail: sbi.nucleodeeditoracao@puc-campinas.edu.br

Apoio:





Trans**Informação**

FICHA CATALOGRÁFICA

Elaborada pelo Sistema de Bibliotecas e
Informação - SBI - PUC-Campinas

Transinformação. Pontifícia Universidade Católica de Campinas. Centro de Ciências Humanas e Sociais Aplicadas. – Campinas, SP, v.1 n.1 (jan./abr. 1989-)

v.26 n.3 set./dez. 2014

Quadrimestral 1989-1999; Semestral 2000-2002; Quadrimestral 2003-
Resumo em Português e Inglês.
ISSN 0103-3786

1. Biblioteconomia – Periódicos. 2. Ciência da Informação – Periódicos.
I. Pontifícia Universidade Católica de Campinas. Centro de Ciências Humanas e Sociais Aplicadas.

CDD 020

Apresentação | *Presentation*

- 225 TransInformação edita número temático - *TransInformação edits thematic issue*
- Rogério Eduardo Rodrigues Bazi
- Seminário Internacional de Bibliometria - *International Seminar on Bibliometrics*
- Jane M. Russell; Maria Victoria Guzmán; Isidro Aguillo; Francisco Collazo Reyes; Rogério Mugnaini

Artigos Originais | *Original Articles*

- 229 Analysis of co-authorship patterns at the individual level
Análise dos padrões de coautoria a nível individual
- Wolfgang Glänzel
- 239 Comunicação científica no Brasil (1998-2012): indexação, crescimento, fluxo e dispersão
Scientific communication in Brazil (1998-2012): Indexing, growth, flow and dispersion
- Rogério Mugnaini; Luciano Antonio Digiampietri; Jesús Pascual Mena-Chalco
- 253 Visibilidad de los estudios en análisis de redes sociales en América del Sur: su evolución y métricas de 1990-2013
Visibility of studies in social network analysis in South America: Its evolution and metrics from 1990 to 2013
- Adilson Luiz Pinto; Audilio Gonzales-Aguilar
- 269 Las prácticas de citación como interpretantes semióticos de acreditación de saberes locales en astronomía: México 1952-1972
Citation practices as semiotic interpretants of certification of local knowledge in astronomy: Mexico 1952-1972
- Francisco Collazo-Reyes; Xochitl Flores-Vargas; Mitzi Lizeth Muñoz-García; Miguel Ángel Pérez-Angón
- 281 Producción científica cubana en Medicina y Salud Pública: Scopus 2003-2011
Cuban scientific production in Medicine and Public Health: Scopus 2003-2011
- Grisel Zacca-González; Benjamín Vargas-Quesada; Zaida Chinchilla-Rodríguez; Félix de Moya-Anegón
- 295 A organização do conhecimento na dinâmica da pesquisa em artigos da literatura científica da Brapci
Knowledge organization in the dynamics of research in the scientific literature of articles at Brapci
- Juliana Lazzarotto Freitas; Bruna Silva do Nascimento; Leilah Santiago Bufrem
- 305 Marketing en el campo de las disciplinas informativas: tendencias de investigación en América Latina (1985-2012)
Marketing in the field of information disciplines: Research trends in Latin America (1985-2012)
- Carlos Luis González-Valiente

- 315 Análisis comparativo de herramientas de recuperación y análisis de información de acceso libre desde una concepción docente
Comparative analysis of information retrieval and analysis of open access tools from an educational concept
• Armando Plasencia-Salgueiro; Bárbara de los Milagros Ballagas-Flores
- 327 Análisis de contenido del dominio tecnológico *vegetable oil combustion*
Content analysis of the technological domain vegetable oil combustion
• Noé Pérez-Arreortúa; Maidelyn Díaz-Pérez; Raudel Giráldez-Reyes; Humberto Andrés Carrillo-Calvet
- 339 Estudio de patentes sobre tecnologías para tratamiento de agua y el agua residual
A study of patenting on water and wastewater treatment technologies
• Odonel González-Cabrera; Taimi Ruiz-Pérez; Mirelys Claro-Pérez; Nancy María Pérez-Pino; Grisel Pérez-Gálvez; Lisbey Collazo-Alfonso
- 349 Tecnologías constituidas, innovaciones en proceso y tecnologías introducidas en el mercado internacional: caso de estudio
Consolidated technologies, ongoing innovations and technologies introduced to the international market: A case study
• Maidelyn Díaz-Pérez; Raudel Giráldez-Reyes; Dayron Armas-Peña; Reinaldo Javier Rodríguez-Font; Elio Atenógenes Villaseñor-García; Humberto Andrés Carrillo-Calvet
- 361 Índice de Autores
Index of Authors
- 363 Índice de Assuntos
Subject Index
- 365 Agradecimientos
Acknowledgements
- 369 Instruções aos Autores
Guide for Authors

TransInformação edita número temático

Nosso periódico TransInformação publica na presente edição textos selecionados do *VII Seminário Internacional de Estudos Quantitativos e Qualitativos da Ciência e Tecnologia*. Apresentamos abaixo os professores parceiros deste número, aos quais reiteramos nossos agradecimentos.

Rogério Eduardo Rodrigues Bazi
Editor

Seminário Internacional de Bibliometria

Como de costume a cada dois anos, realizamos o *Seminário Internacional de Estudos Quantitativos e Qualitativos da Ciência e Tecnologia "Professor Gilberto Sotolongo Aguilar"*, agora em sua sétima edição, sob os auspícios do Congresso Internacional de Informação, realizado em abril último no *Palacio de Convenciones* em Havana, Cuba.

O Seminário de 2014 iniciou-se em 15 de abril, com a inauguração da seção de pôsteres, terminando no dia 17. O programa incluiu 47 apresentações, sendo 28 orais e 19 em forma de pôsteres. Como em anos anteriores, houve um número significativo de estudos de Cuba e México, além de estarem representados outros países, como Bélgica-Hungria, Espanha, Peru e Brasil.

A Chamada de Trabalhos atraiu artigos de pesquisa e de revisão, completos ou em desenvolvimento, bem como estudos de caso, relacionados aos estudos quantitativos e/ou qualitativos da ciência e tecnologia. Estudos Bibliométricos, Cientométricos, Informétricos, Patentométricos e Webométricos foram considerados de particular relevância, sem diminuir a importância de métodos e abordagens analíticos qualitativos.

Apresentações cobriram diversos temas, como *rankings* universitários, estudos de comunicação científica, saúde pública, *marketing* na ciência da informação, astronomia e neurociências, entre outros. De especial interesse foram aqueles relacionados a questões voltadas para a comunidade científica latino-americana, como a qualidade das bases de dados geradas em nível nacional e a representação das publicações nacionais nas bases de dados mais conhecidas, como *Web of Science* e

Scopus. Ambos os aspectos foram considerados significativos para a utilização de fontes tradicionais para medir a produção científica local. Outro tema amplamente comentado foi a queda na produção científica e tecnológica de vários dos países da região, nos últimos três anos.

Como palestrantes convidados, apresentaram-se o Dr. Wolfgang Glänzel e o Dr. Rogério Mugnaini - o primeiro, representando a KU Leuven (Bélgica) e a Biblioteca da Academia de Ciências da Hungria, sediada em Budapeste, com uma palestra intitulada "*Análise dos padrões de coautoria em nível individual*"; e o segundo, representando a Universidade de São Paulo (Brasil), com a palestra "*Comunicação científica no Brasil (1998-2012): indexação, crescimento, fluxo e dispersão*".

Um resultado importante do VII Seminário é este número monográfico da revista *Transinformação*, que contém uma seleção de trabalhos apresentados durante o evento. Embora os Anais do Congresso Internacional de Informação, 2014 incluam pequenos artigos em formato eletrônico sobre as apresentações do Seminário, a publicação de uma versão estendida dos estudos fornece a oportunidade não só para divulgar o Seminário e incentivar uma maior participação no evento, mas também para estender a um público mais amplo o conhecimento acerca da pesquisa atualmente realizada sobre essa temática na América Latina. No mundo de hoje, os estudos de avaliação e prestação de contas são cada vez mais exigidos pelos gestores do sistema de ciência e tecnologia e pelos formuladores de política científica. Por essa razão, a América Latina precisa de um

arcabouço de pesquisa e de pesquisadores dedicados ao avanço do conhecimento nas diferentes especialidades dos estudos métricos da informação, com especial ênfase para as necessidades e características da ciência e tecnologia na região.

Os autores de todos os trabalhos aceitos para apresentação oral no evento foram convidados a apresentar uma versão estendida do seu estudo para possível publicação no periódico *Transinformação*, dando especial atenção aos aspectos teóricos, ao desenvolvimento analítico e à discussão dos resultados. Os manuscritos submetidos foram submetidos a revisores da Revista, resultando em 11 artigos selecionados para compor esta edição especial.

Os dois primeiros estudos apresentam claro contraste nos níveis de agregação de dados. Enquanto o primeiro, de Wolfgang Glänzel, analisou padrões de coautoria de pesquisadores individuais usando uma variedade de métodos bibliométricos e indicadores, o segundo, de Mugnaini e coautores, avaliou grandes quantidade de artigos brasileiros para entender o crescimento, fluxo e dispersão de revistas entre as zonas de Bradford, em um período de 15 anos.

A análise de redes sociais foi utilizada em uma série de estudos, dentre eles o de Pinto e Aguilar, que analisaram a produção latino-americana sobre o tema, entre 1990 e 2013, dando especial atenção aos países, universidades e autores que contribuem nesse sentido.

O estudo de Collazo-Reyes e coautores aplicou uma abordagem qualitativa para analisar as práticas de citação da produção mexicana no campo da astronomia, entre 1952 e 1972. Usando referências como meio de interpretação semiótica, o estudo analisou a relação entre marcas ou sinais associados com filiações locais e sua

relação com os padrões da comunicação científica moderna.

Três estudos focaram temas específicos. Zaccá-González e coautores desenvolveram indicadores de produção, visibilidade e colaboração para a pesquisa em saúde, no período entre 2003 e 2011, usando a base de dados Scopus. Freitas e coautores analisaram conceitos específicos na Organização do Conhecimento, enquanto González-Valiente estudou o papel do *Marketing* no campo das disciplinas de informação.

A edição especial termina com quatro estudos tecnológicos. O primeiro, de Salgueiro e Flores, é de natureza qualitativa e faz uma comparação detalhada de 13 diferentes ferramentas de acesso aberto para recuperação e análise de informação, utilizada em um curso na área da cibernética aplicada. Patentometria, um tema pouco estudado no Brasil, é o foco das outras contribuições: no caso de Arreortúa e coautores, foi aplicada a Combustão de Óleo Vegetal, enquanto Cabrera e coautores estudaram a Tecnologia para Tratamento de Água e Água Residual, e Díaz-Pérez e coautores estudaram o Gás Xisto. Os quatro utilizaram a Classificação Internacional de Patentes, sendo que Salgueiro e Flores, assim como Cabrera e coautores, também usaram análise de redes sociais.

O Seminário, que celebrará sua oitava edição em 2016, continua a ser um local essencial para reunir especialistas da América Latina para apresentar e discutir suas pesquisas sobre temas relevantes para a região.

¹ Jane M. Russell

² Maria Victoria Guzmán

³ Isidro Aguillo

⁴ Francisco Collazo Reyes

⁵ Rogério Mugnaini

¹ Universidad Nacional Autónoma de México. Mexico. *E-mail*: <jrussell@unam.mx>.

² Instituto Finlay. Habana, Cuba. *E-mail*: <mvguzman@finlay.edu.cu>.

³ Consejo Superior de Investigaciones Científicas. Madrid, España. *E-mail*: <isidro.aguillo@cchs.csic.es>.

⁴ Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional. México. *E-mail*: <fcollazo@fis.cinvestav.mx>.

⁵ Universidade de São Paulo, Escola Comunicações e Artes, Departamento de Biblioteconomia e Documentação. São Paulo, SP, Brasil. *E-mail*: <mugnaini@usp.br>.

TransInformação edits thematic issue

The present edition of our periodical *TransInformação* contains texts selected from the *VII International Seminar on Quantitative and Qualitative Studies of Science and Technology*. We present below the professors who participated in this issue and whom we once again thank.

Rogério Eduardo Rodrigues Bazi
Editor

International Seminar on Bibliometrics

As happens every two years in April, the *International Seminar on Quantitative and Qualitative Studies of Science and Technology* "Professor Gilberto Sotolongo Aguilar," now in its seventh edition and held under the auspices of the biennial *International Congress on Information*, took place at the Palacio de Convenciones in Havana, Cuba. The 2014 Seminar began on April 15th with the inauguration of the poster session and ended on April 17th. The program included 47 presentations, 28 oral presentations, and 19 posters. As seen in previous years, there were many studies from Cuba and Mexico, but other countries were also represented such as Belgium-Hungary, Brazil, Peru and Spain.

The Call for Papers attracted research studies, review papers, and case studies completed or in progress related to the quantitative and/or qualitative studies of science and technology. Bibliometric, scientometric, informetric, patentometric and webometric studies were considered to be of particular relevance, without discounting the importance of qualitative analytical methods and approaches.

Presentations covered diverse themes, for example university rankings, scientific communication studies, public health, and marketing in information science, astronomy, and neurosciences, among others. Those related to issues of particular concern were of special interest to Latin American experts, such as quality of national databases and the representation of national publications in the most well-known databases, such as Web of Science and Scopus. Both aspects were

considered significant to the use of mainstream sources for measuring local science. Another widely-commented topic was the decrease in the scientific and technological production of several of the countries in the region in the last three years.

Invited speakers were Dr. Wolfgang Glänzel, KU Leuven, Belgium and Library of the Hungarian Academy of Sciences, Budapest with a talk entitled "Analysis of co-authorship patterns at the individual level" and Dr. Rogério Mugnaini from the University of São Paulo, Brazil, with a presentation called "Scientific communication in Brazil (1998-2012): Indexing, growth, flow and dispersion".

An important outcome of the VII Seminar is this monographic issue of the journal *TransInformação* with a selection of papers presented during the event. Although the proceedings of *International Congress on Information* include short articles in electronic format of the presentations at the Seminar, the publication of an extended version of these studies provides the opportunity not only to publicize the Seminar and encourage greater participation in the event but also for a wider audience to learn of current research in Latin America. In today's world, evaluation and accountability studies are increasingly demanded by research managers and science policy makers. For this reason alone, Latin America needs a body of research and researchers dedicated to advancing knowledge in the different metric specialties of information studies with specific emphasis on the needs and characteristics of science and technology in the region.

Authors of all papers accepted for oral presentation at the event were invited to submit an extended version of their study for possible publication in *Transinformação* giving special attention to theoretical aspects, analytical development and discussion of the results. Submitted manuscripts were externally peer-reviewed by the journal, and 11 articles were selected to make up this special issue.

The first two studies clearly show contrasting levels of data aggregation. While the first by Wolfgang Glänzel looks at co-authorship patterns of individual researchers using a range of bibliometric methods and indicators, the second by Mugnaini and co-authors analyses large data sets of Brazilian articles to understand growth, flow, and dispersion of journals across Bradford zones over a period of 15 years.

Social network analysis was present in a number of studies; among these is the paper by Pinto and Aguilar who analysed Latin American production on this subject from 1990 to 2013, giving special attention to contributing countries, universities and authors.

The study by Collazo-Reyes and co-authors used a qualitative approach to analyse the citation practices of the Mexican production in the field of astronomy from 1952 to 1972. Using references as means for semiotic interpretation, the study looked at the relationship between marks or signs associated with local affiliations and that of modern scientific communication patterns.

Three studies focussed on specific topics. Zacca-González and co-authors developed production, visibility,

and collaboration indicators for health research for the period from 2003 to 2011 using the Scopus database. Freitas and co-authors analysed specific concepts in knowledge organization while González-Valiente studied marketing in the field of information disciplines.

The special issue closes with four technological studies. The first by Salguero and Flores is qualitative in nature and makes a detailed comparison of 13 different open access tools for the retrieval and analysis of information used in a course in the area of Applied Cybernetics. Patentometrics, a subject little studied in Brazil, is the focus of the other contributions; in the case of Arreortúa and co-authors it is applied to vegetable oil combustion; Cabrera and co-authors study water and wastewater treatment technology; and Díaz-Pérez and co-authors study shale gas. All four use the International Patent Classification; Salguero and Flores, and Cabrera and co-authors also use social network analysis.

The Seminar which will celebrate its eighth edition in 2016 continues to be an essential venue to bring together specialists from Latin America to present and discuss their research on relevant topics for the region.

¹Jane M. Russell

²Maria Victoria Guzmán

³Isidro Aguillo

⁴Francisco Collazo Reyes

⁵Rogério Mugnaini

¹ Universidad Nacional Autónoma de México, Mexico. *E-mail*: <jrussell@unam.mx>.

² Instituto Finlay, Habana, Cuba. *E-mail*: <mvguzman@finlay.edu.cu>.

³ Consejo Superior de Investigaciones Científicas, Madrid, España. *E-mail*: <isidro.aguillo@cchs.csic.es>.

⁴ Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, México. *E-mail*: <fcollazo@fis.cinvestav.mx>.

⁵ Universidade de São Paulo, Escola Comunicações e Artes, Departamento de Biblioteconomia e Documentação, São Paulo, SP, Brasil. *E-mail*: <mugnaini@usp.br>.

Analysis of co-authorship patterns at the individual level¹

Análise dos padrões de coautoria a nível individual

Wolfgang GLÄNZEL^{2,3}

Abstract

Publication activity, citation impact and communication patterns, in general, change in the course of a scientist's career. Mobility and radical changes in a scientist's research environment or profile are among the most spectacular factors that have effect on individual collaboration patterns. Although bibliometrics at this level should be applied with the utmost care, characteristic patterns of an individual scientist's research collaboration and changes in these in the course of a career can be well depicted using bibliometric methods. A wide variety of indicators and network tools are chosen to follow up the evolution and to visualise and to quantify collaboration and performance profiles of individual researchers. These methods are, however, designed to supplement expert-opinion based assessment and other qualitative assessments, and should not be used as stand-alone evaluation tools. This study presents part of the results published in an earlier study by Zhang and Glänzel (2012)⁴ as well as new applications of these methods.

Keywords: Bibliometrics indicators. Bibliometry. Scientific production.

Resumo

Os padrões da atividade de publicação, do impacto da citação e comunicação, em geral, mudam ao longo da carreira de um cientista. A mobilidade e as mudanças radicais no ambiente de pesquisa de um cientista ou no seu perfil estão entre os fatores mais espetaculares que têm efeito sobre os padrões de colaboração individuais. Embora a bibliometria, neste nível, deva ser aplicada com o máximo de cuidado, os padrões característicos da colaboração e as mudanças na pesquisa de um cientista individual no curso de sua carreira podem ser bem representados pelos métodos bibliométricos. Uma grande variedade de indicadores e as ferramentas de rede são utilizadas para acompanhar a evolução, visualizar e quantificar o perfil de colaboração e desempenho de pesquisadores individuais. Estes métodos são, no entanto, concebidos para complementar a avaliação baseada na opinião do especialista e em outras avaliações qualitativas, e não devem ser usadas como os únicos instrumentos de avaliação. Este estudo apresenta parte dos resultados publicados em um estudo anterior por Zhang e Glänzel, (2012)⁴ bem como novas aplicações desses métodos.

Palavras-chave: Indicadores bibliométricos. Bibliometria. Produção científica.

Introduction

The evolution from "little scientometrics" to "big scientometrics" (Glänzel & Schoepflin, 1994) is characterised by two cardinal signs (Glänzel & Wouters,

2013): In the last quarter of the 20th century, bibliometrics evolved from a sub-discipline of library and information science to an instrument for research evaluation and benchmarking called "perspective shift" (Glänzel *et al.*, 2006; Wouters, 2014). As a consequence of this

¹ Study presented at the VII International Seminar on Quantitative and Qualitative Studies on Science and Technology "Professor Gilberto Sotolongo Aguilar" at the XIII International Congress of Information - INFO' 2014. Habana, Cuba.

² KU Leuven, Centre for Research & Development Monitoring, Department of Managerial economics, Strategy and Innovation. Waaistraat 6, 3000, Leuven, Belgium. *E-mail:* <wolfgang.glanzel@kuleuven.be>.

³ Library of the Hungarian Academy of Sciences, Department of Science Policy & Scientometrics. Budapest, Hungary.

⁴ Figures 1-3 of the present paper are reproduced from Zhang and Glänzel (2012) with permission of the publisher.

Received in 6/27/2014 and approved in 9/9/2014.

perspective shift, new fields of applications and challenges have opened to bibliometrics, although many tools continued to be designed for use in the context of scientific information, information retrieval and librarianship. In other words, these became used in a context for which they were not designed (the Journal Impact Factor). Secondly, due to the dynamics in the evaluation of research, focus has shifted away from macro studies towards meso and micro studies of both actors and topics. More recently, the evaluation of research teams and individual scientists has become a central issue in services based on bibliometric data.

The rapid development of information technology has opened bibliometrics to a broader audience. Passive “consumers” in science policy, research management and the scientific community as well as active users and “semi-professionals” producing bibliometric indicators for various purposes have gained access to the necessary data and tools. Above all, electronic communication and the Web have paved the way for some type of democratisation of bibliometrics resulting in a rather vulgar version of democracy with anarchistic features (Glänzel & Hornbostel, 2011). Thus, bibliometrics has become available to practically any user, notably at the micro level.

While bibliometric macro and meso data still preserve a certain extent of anonymity, micro-level data call a spade a spade. Researchers have thus become more susceptible to the consequences of bibliometric practice since they are increasingly concerned by policy use and misuse of bibliometric methods (Glänzel & Debackere, 2003). Sometimes they even feel they are victims of the evaluation. Bibliometric techniques should therefore always be used in a proper context, notably in combination with “qualitative methods” and special caution is always called for at this level.

Recently Glänzel and Wouters (2013) formulated 10 recommendations for bibliometrics “The dos and don’ts in individual-level bibliometrics”. In particular, Glänzel and Wouters recommended the use of individual level bibliometrics always based on the particular research portfolio of the relevant researcher. The best method to do this may be the design of individual

researchers’ profiles combining bibliometrics with qualitative information about their careers and working contexts.

As regards the quantitative component of research assessment at this level, bibliometrics can be used to zoom in on a scientist’s career. Here the evolution of publication activity, citation impact, mobility and changing collaboration patterns can be monitored. It is not easy to quantify the observations and the purpose is not to build indicators for possible comparison, but to use bibliometric data to visually and numerically depict important aspects of the progress of a scientist’s career. In the following section, I will focus on scientists’ publication activity, their co-authorship patterns and the citation impact at different stages of their career.

Methods

Although bibliometrics at the level of individuals should be applied with the utmost caution, characteristic patterns in a scientist’s career can be well depicted with bibliometric methods. These methods refer to the following topics.

- Communication patterns, in general, publication activity and citation impact, in particular, change in the course of a scientist’s career.

- Mobility, promotion or a change in a scientist’s research environment, usually results in structural changes of collaboration patterns as well.

According to the recommendations by Glänzel and Wouters (2013), the combination of bibliometrics with career analysis is one of the opportunities of quantitative science studies at the individual level. This, of course, requires assessment on the basis of a scientist’s complete *oeuvre*. In this context bibliometrics can be used to zoom in on various stages and phases in a scientist’s career. Here the evolution of publication activity, citation impact, mobility and changing collaboration patterns can be monitored, as has previously been mentioned. The first results were recently published by Zhang and Glänzel (2012). Here I will present several examples that can be used to analyse productivity and impact patterns at different stages of a

career. Some of these examples are taken from the aforementioned study by Zhang and Glänzel, others have been prepared for the present paper. In particular, the following five issues will be analysed.

- Evolution of publication activity and citation impact in the scientist's life cycle;
- Topicality of highly cited papers;
- Evolution of the number of co-authors in the course of a scientist's career;
- Partnership ability and role of co-authorship in productivity and citation impact;
- The scientist's position and role in his/her collaboration network.

Of course, bibliometrics alone cannot answer all questions concerning the performance of individual scientists, but it can already provide a valuable indication that allows deeper analysis with the help of qualitative information about careers and working contexts, as was mentioned in the introduction.

All methods presented in this section will be illustrated by examples. Authors in several research fields appearing in these examples will, however, be treated anonymously.

Bibliometric career analysis of individual scientists

Cronin and Meho (2007) have previously pointed out the close relationship between creativity and age in the field of information science. The easiest way to show this is certainly the application of *age pyramids*. This idea goes back to demographics, where the population structure, composition and age of the human population is quantitatively described. The population pyramid is actually an elementary tool to reflect the age structure of a given population. In demographics the age distribution in a human population is plotted in a double bar diagram, in particular, male age groups are plotted against the corresponding female groups. Usually about 5-7 paradigmatic shapes are distinguished. These reflect different paradigmatic types of growth characteristics of a given population. Here we focus just on three typical shapes, namely:

- triangle, pagoda and bell shape (three cases of growth patterns with high fertility but different extent of infant mortality);
- beehive shape (stands for stationary structure with low infant mortality);
- "onion" shape (reflects superannuation of the population).

Analogously, "age pyramids" using double bar diagrams of publication activity (at the given time period) and citation impact (based on citations received in that time period to all previously published papers) - instead of juxtaposing gender-related age groups in human population - can be used to reflect important changes in the course of a scientist's career (Zhang & Glänzel, 2012). The authors have pointed out essential differences between the original demographic and "scientometric" age-pyramid model. While in population, data bars representing male and female groups usually follow the same basic shape, in the bibliometric case the shapes for publications and citations might distinctly differ. For instance, triangle-shaped productivity might be contrasted by an onion-shaped citation impact. In this context I would like to stress that bibliometric age pyramids reflect both subject specific peculiarities and individual "performance" patterns. This is shown using the example of four selected scientists already introduced by Zhang and Glänzel (2012). These scientists stand for four different research areas, particularly, the life sciences, natural sciences, mathematics and the social sciences. Due to subject-specific biases the shape of the age distribution in the natural sciences and in mathematics is expected to be "flatter" in contrast with more skewed distributions in life sciences and social sciences. Moreover, in these general patterns we find interesting individual characteristics in the pyramid shapes of the four authors (Figure 1).

While the beehive in the case of the citation impact of the second author and the onion in the third case generally mirror the corresponding shapes of publication age, the onion shape of publication age is contrasted by a citation triangle in the first case. The patterns for the latter authors reflect a steady growth of both productivity and impact. Some reasons for the deviation of impact from productivity patterns have been

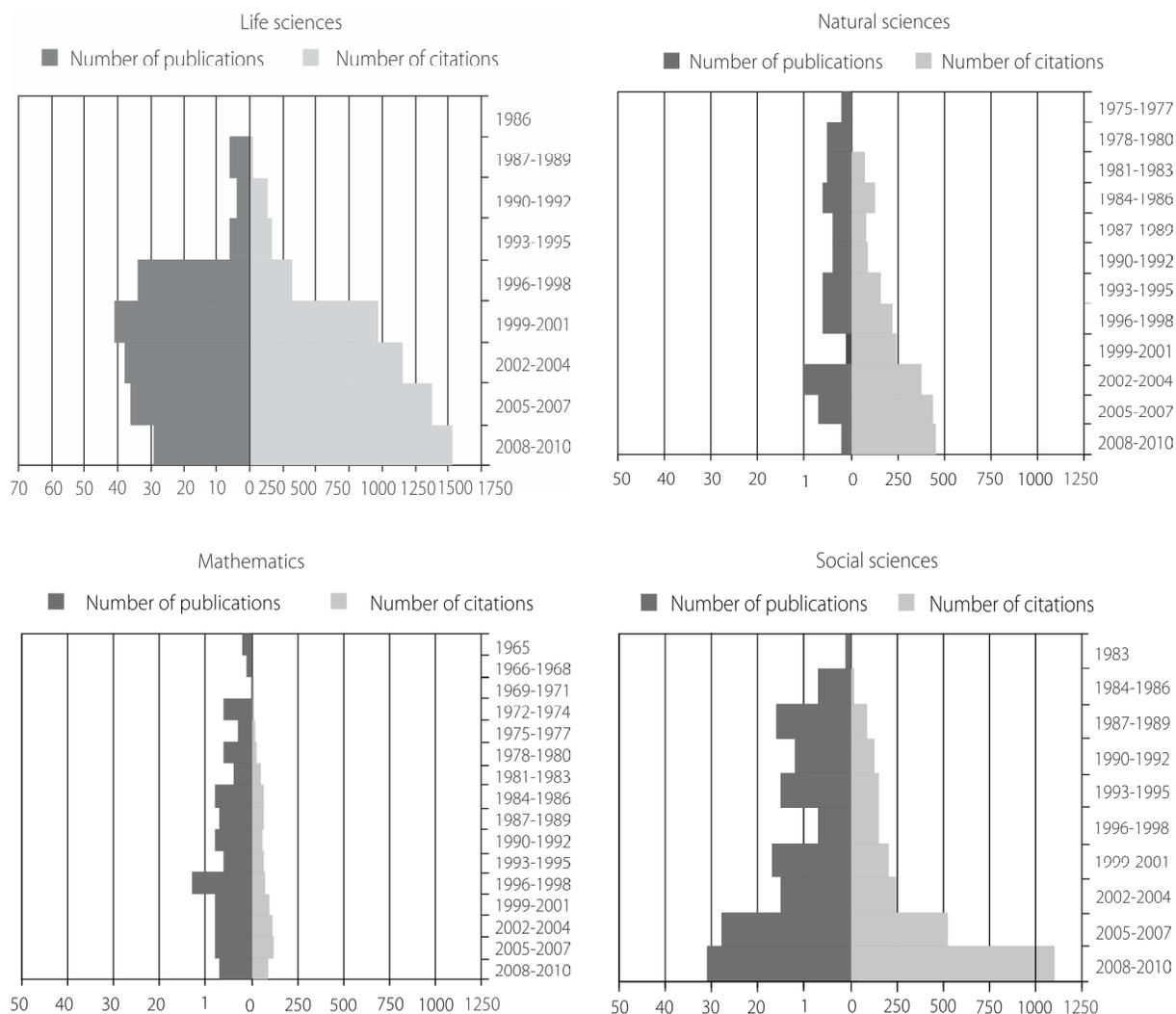


Figure 1. Scientometric age pyramids for four scientists according to Zhang and Glänzel (2012).

Note: The x-axis displays the number of publications and citations, respectively.

discussed by Zhang and Glänzel. One should bear in mind that citation impact refers to the present *and* the past, while productivity always reflects the situation of the period under study. Further analysis and interpretation of these shapes might therefore reveal details on the relationship between the impact of recent and former research. This can be deepened by analysing the constitution of an author's highly cited papers over time. This idea goes back to the h-index sequence proposed by Liang (2006) to measure the dynamics of the h-index in a scientist's career. This idea has been extended by Zhang and Glänzel (2012) to the mean age of publications of the h-core sequence, where the h-core

sequence is defined analogously to the h-core for the h-index sequence. The calculation of the mean age sequence of the h-core follows the algorithm proposed by Zhang and Glänzel (2012):

- The h-core is formed by those papers that have received at least h citations, where h denote the actual value of an h-index.

- The h-core sequence: we first calculated the h-index for papers published in the first year of their career, then the first two years, the first three years, and so on until the most recent year is reached.

- The mean age of publications of this h-core sequence is calculated, which expresses whether the

more recent or the older publications are predominant in the respective h-core.

According to the above-mentioned study by Zhang and Glänzel (2012) we can distinguish four paradigmatic patterns with case 1 representing the standard situation. A convex shape stands for "superannuation" of citation impact (mainly old papers are cited), while a concave age curve reflects an increasing number of recent papers entering the h-core.

- The mean age of the h-cores follow a linear function of time. This reflects a steady growth of the age of highly-cited papers.

- A convex shape reflects accelerated growing age of highly-cited papers. In verbal terms, most cited papers have rather been published in earlier stages of the scientist's career.

- A concave shape reflects decreasing age of highly-cited papers. This is the opposite situation to the previous case, among the most cited papers one finds more recent ones.

- An "indefinite" shape covering all cases not listed above.

The mean age sequences of the h-core for the same four scientists in the "demographic" representation are plotted in Figure 2. The shape for authors #2 might be considered to be in line with the standard (linear curve), while the age sequence of author #4 corresponds to the concave case. The fluctuations in the 1970s (author #3) bear witness to quite dramatic changes in the constitution of the scientist's h-core. Publication activity, citation impact and their change in time are, however, not independent of collaboration and team work.

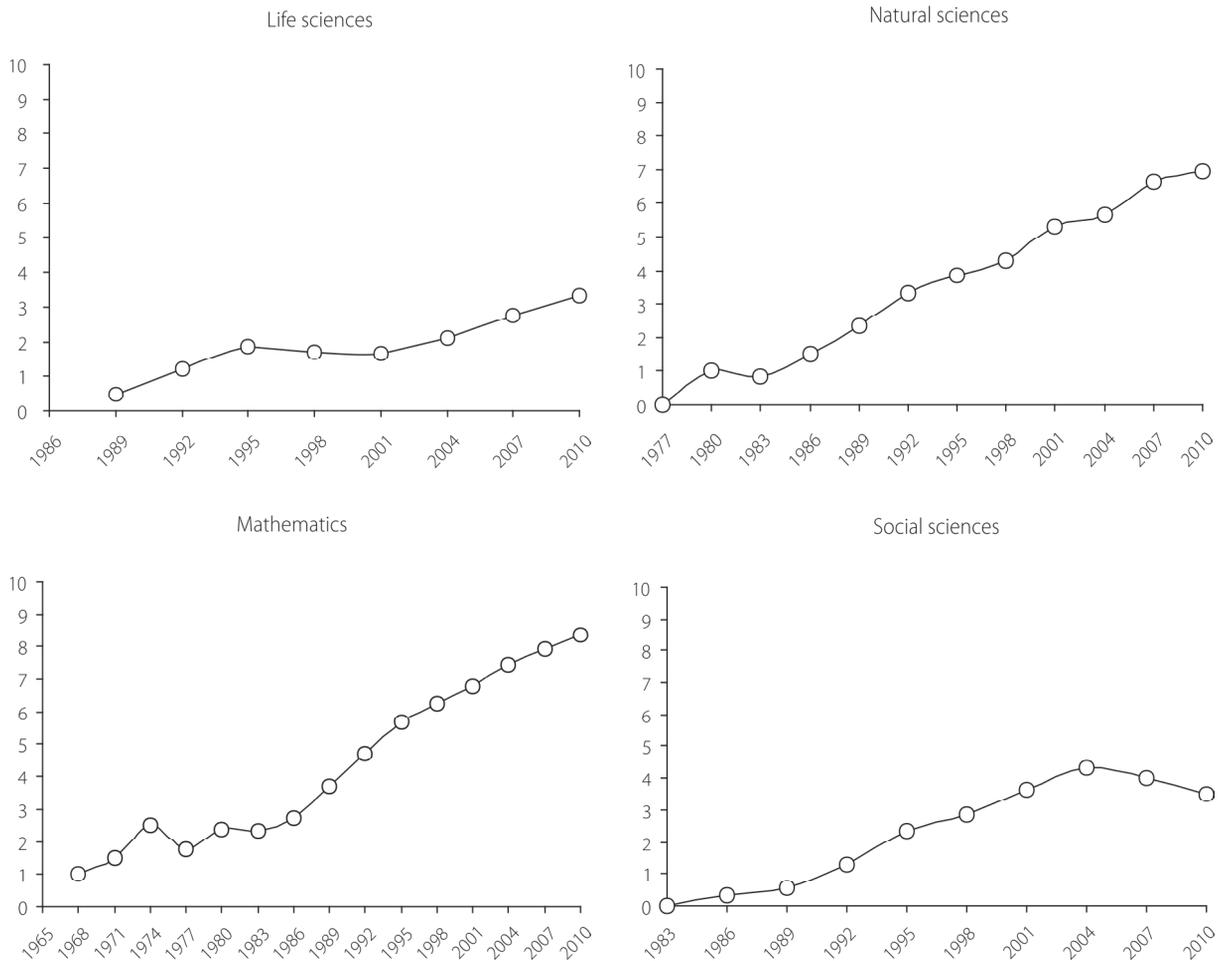


Figure 2. Mean age sequence of the h-core displayed on the y-axis for four scientists according to Zhang and Glänzel (2012).

Number and role of co-authors, frequency of co-authorship might strongly affect bibliometric indicators. A logical consequence is therefore the extension of individual-level bibliometrics to the analysis of collaboration patterns. This will be done in the following subsection.

Co-authorship patterns of individual scientists

In order to quantify the above-mentioned connectedness, Zhang and Glänzel (2012) have analysed “the extent of co-authorship”, which denotes the number of different co-authors of the scientist under study. The objective was to analyse whether the changing size of

cooperation has any influence on the authors’ productivity. A positive relation between collaboration and productivity has been found in early scientometric studies: Beaver and Rosen (1979) concluded that collaboration is associated with higher productivity. This finding has been reconsidered by Braun *et al.* (2001) and Glänzel (2002), who found that increasing co-operativity goes only to a certain extent with higher publication activity and beyond some subject-specific threshold co-operativity turns into a “negative” effect in terms of productivity. In addition, the generally positive effect of collaboration on citation impact in practically all subject fields and at all levels of aggregation has been shown (Narin & Whitlow, 1990; Moed *et al.*, 1991; Narin *et al.*, 1991). This consequently raises some questions, namely:

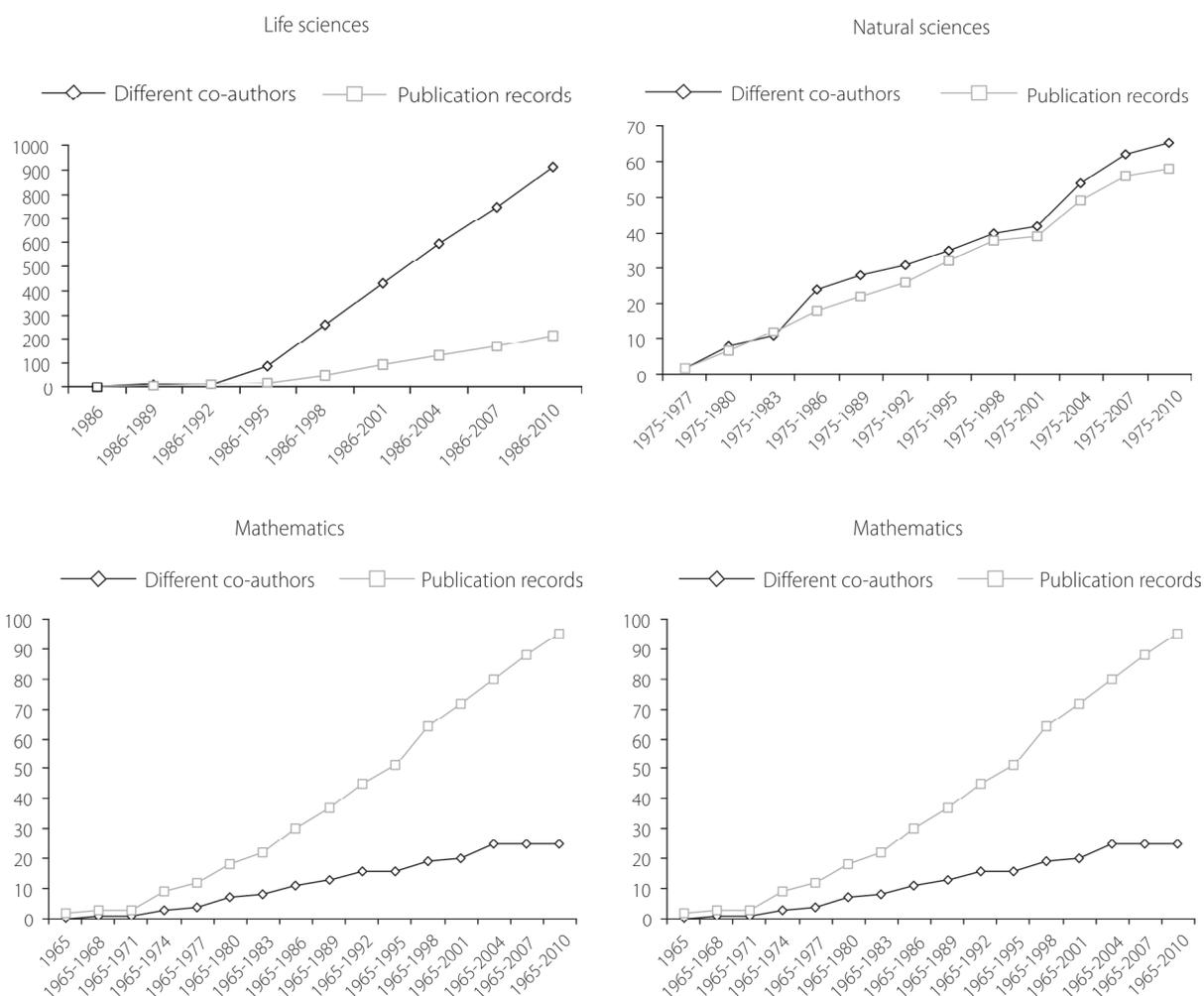


Figure 3. Co-author and publication sequence for four scientists according to Zhang and Glänzel (2012).

how does the co-operativity of individual scientists changes with their productivity in the course of their career? and is there any positive or negative effect of intensive, stable or occasional collaboration links on productivity and possibly on citation impact as well? "The extent of co-authorship" can be expressed by the number of different co-authors a scientist has in a given period. Figure 3 shows the cumulative number of (different) co-authors and publications again for the same four scientists as above. Furthermore, in this context subject-specific effects are visible; for instance, co-authorship of the mathematician seems to be limited by the smaller community as compared to the scientist in the life sciences. Thus the scientists in life sciences and natural sciences, in general, had more co-authors than publications, and the cases of mathematics and social sciences show the opposite picture (Zhang & Glänzel, 2012).

More recently, some new indexes have been proposed to quantify effects of co-authorship in the context of productivity and citation impact. The first one was proposed by Hirsch (2010). He proposed a new index called ("non-selfconsistent") ' h ' to characterize the scientific output of a researcher, which takes into account the effect of multiple co-authorship. According to its definition *A paper belongs to the ' h core of a scientist if it has $\geq h$ citations and in addition belongs to the h -core of each of the co-authors of the paper.*

Schubert (2012) goes a step further. He defined a Hirsch-type index to characterise "partnership" in scientific publication output: *An actor is said to have a partnership ability index φ , if with φ of his/her n partners had at least φ joint actions each, and with the other $(n - \varphi)$ partners had no more than φ joint actions each.*

Similarly to the h -index, which combines publication activity with citation impact, the φ -index combines two important features: publication activity with the frequency of joint activity. Some basic properties of the φ index are listed below.

- $\varphi = 0$: The author has only single-authored papers.

- $\varphi = 1$: Three cases are possible.

1. The author has only double-authored papers with the very same co-author each (*monogamy*).

2. If the author had an arbitrary number of co-authored papers with no co-authors occurring more than once (*total promiscuity*).

3. If the author had an arbitrary number of double-authored papers with the same co-author and an arbitrary number of co-authored papers with other authors such that no co-author occurs more than once (Rousseau, 2012).

- $\varphi > 1$: In all other cases.

According to Schubert, low values reflect a scanty or inconsistent set of co-authors, while high values reflect a wide and persistent co-authorship network.

After having recalled these new measures, we will use Hirsch-type indices along with egocentric networks to shed some light on collaboration patterns in a scientist's career. The aim is to supplement bibliometric indicators at this level, in order to show the extent to which the author's performance is related to his/her own and the colleagues' activity, position and impact. In order to analyse scientists' position among their collaborators and co-authors the following set of indicators is used.

- Number of papers and h -index

- Number and share of single authored papers in all papers

- Number and share of single authored papers in the h -core

- Share of h -core co-authors

- φ index

In order to illustrate the analysis, 13 collaborating authors were chosen as an example. The authors have European affiliation and are active in information science. Their physical age ranges between about 35 to 80 years. Again authors are anonymised and this time denoted by the capital letters A - M, where author "A" is chosen for the egocentric network model. Using the above-mentioned indicators in conjunction with network analysis, among others, the following questions can be answered.

- Do authors preferably work alone, in stable teams, or do they rather prefer occasional collaboration?

- Who are the collaborators and are the scientists rather 'junior', 'peers' or 'senior' partners in these relationships?

The question of who is collaborating with whom in scientific research - and are there general rules for co-operation of authors of similar/different academic age and position - has been studied in bibliometrics for a long time. Kretschmer (1994), for instance, has analysed aspects of social stratification in scientific collaboration. She found that extramural collaboration is rather characterised by similarity of the social status whereas intramural collaboration shows significant differences of the social status of the co-authors. She called these effects "Birds of a feather flock together" and "Opposites attract" and observed that both are frequent in an academic career and rather depend on the nature of collaboration. In a more recent study, Hu *et al.* (2013) attempted to answer the question of whether scientists of young academic age prefer collaborating with older ones and vice versa. They concluded on the basis of their observations that age, in general, does not play a determinant role. Thus there is no general answer and co-authorship links need to be analysed individually indeed. This will be done using the above example. Answering the above questions might help understand the scientist's own role and position in his/her research environment.

The indicator values for seven of the 13 authors can be found in Table 1. Since all authors are collaborators to a certain extent, one cannot expect serious subject-based biases across their profiles. All characteristics are therefore mainly due to performance, (academic) age and position. Authors "A", "E" and "B" are clearly the "seniors" among the selected collaborating scientists. The number

and share of co-authors as well as their share in the h-core provide important information about the co-authors' role in producing research output and in highly cited papers, in particular.

Comparison of the corresponding indicator values of author "E" and "J" allows the conclusion that above all the co-authors of "J" are responsible for "J's" high-impact papers. Usually about 50% of the co-authors contribute to high-impact papers. This alone does not point to continuous research in stable teams. The remarkably large number of "E-type" co-authors, which even exceeds the number of his papers, might only serve as a counter example. Collaboration with "E" is, however, not merely occasional as his large ϕ -value index substantiates. There is another remarkable detail: ϕ -values do not necessarily correlate with the number of co-authors. The comparison of "A's" large ϕ -index with those of "B" and "G" shows that a higher index value might be associated with a lower number of co-authors.

The above observations can be deepened by the analysis of bilateral collaboration links. Figure 4 shows the egocentric network from the viewpoint of author "A". Here all 13 selected co-authors are displayed. The size of the circles is proportional to the corresponding scientists' publication output; the thickness of lines corresponds to the strength of co-authorship links. According to publication output, "B" and "M" can therefore be considered 'peers' with respect to "A", while "D", "F", "H", "I", "J", "K", "L" are his "junior" collaborators. By contrast, "E" and "C" can be considered "seniors". Strong links with "juniors" often point to the role of supervisor, and indeed, "D", "H"

Table 1. Hirsch-type indicators to characterise co-authorship.

Author	N	h	single (all) count	%	single (h-core) count	%	co-authors			ϕ
							all	h-core	%	
A	200	32	34	17.0	6	18.8	27	14	51.9	9
B	167	30	45	26.9	1	3.3	52	25	48.1	5
D	34	11	0	0.0	0	0.0	16	9	56.3	4
E	418	44	1	0.0	0	0.0	563	169	30.0	15
G	66	19	3	4.5	1	5.3	53	25	47.2	6
J	18	8	4	22.2	0	0.0	10	7	70.0	2
K	8	4	1	12.5	0	0.0	3	3	100.0	2

Note: Data sourced from Thomson Reuters Web of Knowledge and retrieved in September 2013.

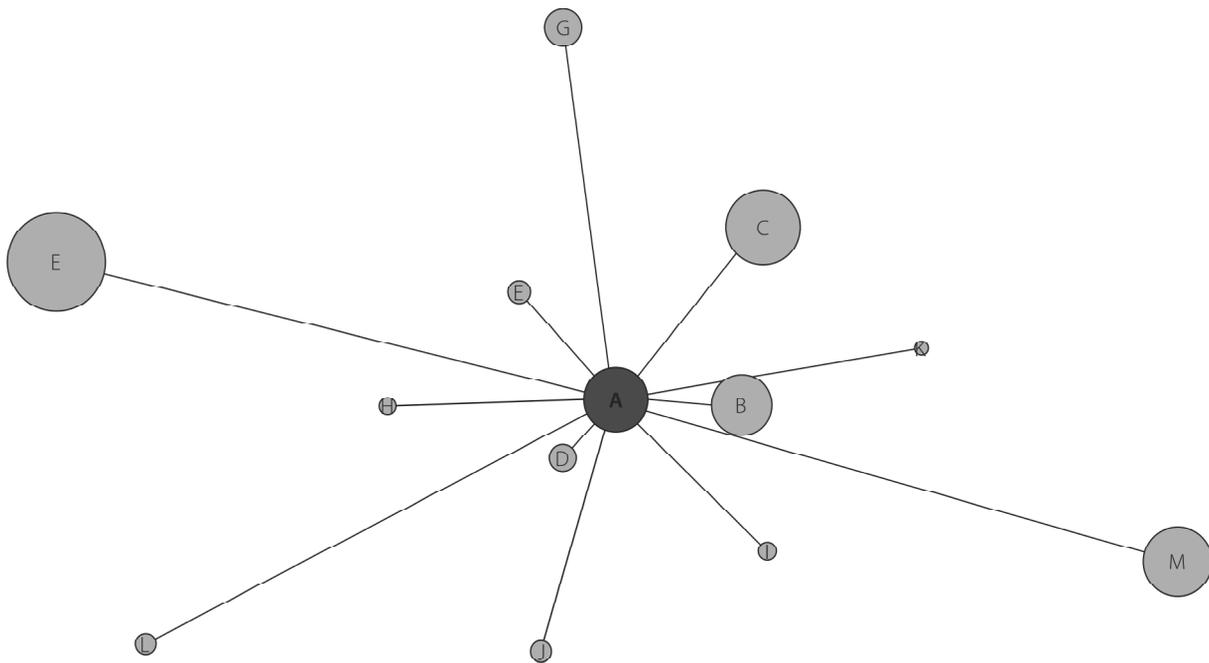


Figure 4. Egocentric network of an author allowing conclusions on the role of co-authors using Pajek with Kamada-Kawai layout.

Note: Data sourced from Thomson Reuters Web of Knowledge and retrieved in September 2013.

and “K” were PhD students of “A”. The strong link with peer “B”, however, reflects long-time collaboration in a stable team and the weak link with “senior” “D”, finally, reveals occasional co-author relationship.

Both exercises proposed in this subsection, namely the co-authorship-related indicators and the network analysis, provide details that complement the “demographic” indicators described in Section *Bibliometric career analysis of individual scientists* by shedding light on the scientist’s position in the network of scholarly communication. A dynamic approach to capture the evolution of the scientist’s position is also possible.

Conclusion

Bibliometric indicators and network analysis provide valuable information on the performance of individual scientists. However, this information should be considered supplementary. In individual research assessment, the emphasis should always be laid on ‘qualitative’ methods. In individual-level evaluation, the added value of bibliometrics depends on how and in

which context bibliometrics is applied. The advantage of obtaining scores and numerical values, of repeating the exercise years later using the same methods, and of having the opportunity of monitoring the change in values over time should, of course, not be underestimated. Bibliometric in-depth analysis of the *evolution of publication activity and citation impact* in the course of a scientist’s career can also help interpret bibliometric standard indicators at this level.

Co-authorship analysis can be used to determine the position of an author in the collaboration network and might provide important information on the scientists’ own contribution to the research output reported in their Curriculum Vitae. In conjunction with the h-core analysis, this reveals details on the extent of the scientist’s real contribution to his/her research output and the citation impact these publications have achieved.

Acknowledgement

The author would like to thank Sarah Heeffer, KU Leuven, for her contribution to collecting, cleaning and processing the underlying bibliographic data.

References

- Beaver, D.D.; Rosen, R. Studies in scientific collaboration: Part II: Scientific co-authorship, research productivity and visibility in the French elite. *Scientometrics*, v.1, n.2, 133-149, 1979.
- Braun, T.; Glänzel, W.; Schubert, A. Publication and co-operation patterns of the authors of neuroscience journals. *Scientometrics*, v.51 n.3, p.499-510, 2001.
- Cronin, B.; Meho, L. Timelines of creativity: A study of intellectual innovators in information science. *Journal of the American Society for Information Science and Technology*, v.58, n.13, 1948-1959, 2007.
- Glänzel, W.; Schoepflin, U. Little scientometrics, big scientometrics - and beyond. *Scientometrics*, v.30 n.2-3, p.375-386, 1994.
- Glänzel, W. Co-authorship patterns and trends in the sciences (1980-1998): A bibliometric study with implications for database indexing and search strategies. *Library Trends*, v.50, n.3, p.461-473, 2002.
- Glänzel, W.; Debackere, K. On the opportunities and limitations in using bibliometric indicators in a policy relevant context. In: Ball, R. (Ed.). *Bibliometric analysis in science and research: Applications, benefits and limitations*. Zürich: Publication Forschungszentrum, 2003. p.225-236.
- Glänzel, W. *et al.* A concise review on the role of author self-citations in information science, bibliometrics and science policy. *Scientometrics*, v.67, n.2, p.263-277, 2006.
- Glänzel, W.; Hornbostel, S. *History and institutionalisation of scientometrics*. Vienna: European Summer School for Scientometrics, 2011.
- Glänzel, W.; Wouters, P. *The dos and don'ts in individual-level bibliometrics*. In: International Conference on Scientometrics and Informetrics, 14., 2013, Vienna. *Electronic proceedings...* Vienna: ISSI, 2013. v.1. Available from: <http://www.issi2013.org/Images/ISSI_Proceedings_Volume_1.pdf>. Cited: May, 2014.
- Hirsch, J.E. An index to quantify an individual's scientific research output that takes into account the effect of multiple co-authorship. *Scientometrics*, v.85, n.3, p.741-754, 2010.
- Hu, Z. *et al.* Do new scientists prefer collaborating with old scientists? And vice versa? In: International Conference on Scientometrics and Informetrics, 14., 2013, Vienna. *Electronic proceedings...* Vienna: ISSI, 2013. v.2. Available from: <http://www.issi2013.org/Images/ISSI_Proceedings_Volume_2.pdf>. Cited: May, 2014.
- Kretschmer, H. Co-authorship networks of invisible colleges and institutional communities. *Scientometrics*, v.30, n.1, p.363-369, 1994.
- Liang, L. H-index sequence and h-index matrix: Constructions and applications. *Scientometrics*, v.69, n.1, p.153-159, 2006.
- Moed, H.F. *et al.* International scientific co-operation and awareness within the European Community: Problems and perspectives. *Scientometrics*, v.21 n.3, p.291-311, 1991.
- Narin, F.; Stevens, K.; Whitlow, E.S. Scientific co-operation in Europe and the citation of multinationally authored papers. *Scientometrics*, v.21, n.3, p.313-323, 1991.
- Narin, F.; Whitlow, E.S. *Measurement of scientific co-operation and co-authorship in CEC-related areas of science*. v. 1-2, Brussels: Commission of the European Communities, 1990.
- Rousseau, R. Comments on a Hirsch-type index of co-author partnership ability. *Scientometrics*, v.91, n.1, p.309-310, 2012.
- Schubert, A. A Hirsch-type index of co-author partnership ability. *Scientometrics*, v.91, n.1, p.303-308, 2012.
- Wouters, P. *The citation: From culture to infrastructure*. In: Blaise, C. & Cassidy, S. (Ed.). *Next generation metrics: Harnessing multidimensional indicators of scholarly performance*. Cambridge: MIT Press, 2014. p.76-116.
- Zhang, L.; Glänzel, W. Where demographics meets scientometrics: Towards a dynamic career analysis. *Scientometrics*, v.91, n.2, p.617-630, 2012.

Comunicação científica no Brasil (1998-2012): indexação, crescimento, fluxo e dispersão¹

Scientific communication in Brazil (1998-2012): Indexing, growth, flow and dispersion

Rogério MUGNAINI²

Luciano Antonio DIGIAMPIETRI³

Jesús Pascual MENA-CHALCO⁴

Resumo

A produção científica brasileira vem demonstrando ascensão no cenário mundial, o que pode decorrer da internacionalização da ciência nacional e/ou das revistas nacionais. Nesse contexto, este estudo reuniu a produção científica brasileira das bases *Web of Science* e *Scientific Electronic Library Online* entre 1998 e 2012, com o fim de analisar como a internacionalização tem influenciado a performance do país. Com base em Zonas de Bradford da produção das diversas áreas, buscou avaliar o aumento do número de revistas em cada uma das zonas, assim como a proporção de artigos publicados em revistas nacionais, ao longo de cinco triênios. Observou-se que o número de revistas indexadas exclusivamente na *Scientific Electronic Library Online* cresceu no período, com exceção do triênio 4, quando aproximadamente 60% da produção nacional foi publicada em revistas estrangeiras (*Web of Science*). Notou-se que o aumento mais generalizado do número de revistas das áreas deu-se primeiramente na zona 3 (triênio 2004-2006), depois na zona 2 (triênio 2007-2009) e finalmente na zona 1 (triênio 2010-2012). Já o percentual da produção em revistas nacionais diferenciou três grupos de áreas: ciências sociais e humanas, onde as zonas 1 e 2 são exclusivas de revistas nacionais, e a zona 3 vem dando lugar a publicações estrangeiras; Física, Ciência do Espaço e Imunologia, cuja produção é exclusivamente internacional; e as demais áreas, de ciências exatas e da terra, saúde e biológicas, que vêm aumentando o percentual nas zonas 1 e 2, permanecendo a zona 3 direcionada a publicações internacionais.

Palavras-chave: Avaliação de pesquisa. Brasil. Produção científica. Revistas científicas nacionais. Zonas de Bradford.

Abstract

The Brazilian scientific production growth in the international scenario may be due to the internationalization of Brazilian science and journals. Within this context, this study collected Brazilian scientific production from the Web of Science and Scientific Electronic Library Online databases - data from 1998 to 2012 - in order to analyze how internationalization influenced the performance of the country. Based on the Bradford Zones and considering the different fields, the aim of this study is to assess the growth in the number of journals in each zone, as well as the proportion of the papers published in national journals throughout five triennia. It was possible to observe that the number of journals exclusively indexed in Scientific Electronic Library Online increased in the period, except in the fourth triennium when several journals were also indexed in the Web of Science. However, about 60% of the national production is published in international journals (Web of Science). The most generalized growth in the number of journals from different fields occurred initially in zone 3 (triennium

¹ Trabalho apresentado no VII Seminário Internacional sobre Estudos Quantitativos e Qualitativos da Ciência e Tecnologia "Prof. Gilberto Sotolongo Aguilar" no XIII Congresso Internacional de Informação - INFO'2014. Habana, Cuba.

² Universidade de São Paulo, Escola Comunicações e Artes, Departamento de Biblioteconomia e Documentação. Av. Prof. Lúcio M. Rodrigues, 443, 05508-020, São Paulo, SP, Brasil. Correspondência para/Correspondence to: R. MUGNAINI. E-mail: <mugnaini@usp.br>.

³ Universidade de São Paulo, Escola de Artes, Ciências e Humanidades. São Paulo, SP, Brasil.

⁴ Universidade Federal do ABC, Centro de Matemática, Computação e Cognição. Santo André, SP, Brasil.

Recebido em 3/6/2014, reapresentado em 11/8/2014 e aceito para publicação em 1/9/2014.

2004-2006), then in zone 2 (triennium 2007-2009), and finally in zone 1 (triennium 2010-2012). On the other hand, the percentage of production in national journals across the fields presented three different behaviors: in Social and Human Sciences zones 1 and 2 contains exclusively national journals, and zone 3 is receiving international journals; in Physics, Space Science and Immunology production is exclusively international; in other fields such as Exact & Earth, Health and Biological Sciences there is an increasing percentage of production in national journals in zone 1 and 2, but zone 3 remains international.

Keywords: Research evaluation. Brazil. Scientific production. National scientific journals. Bradford zones.

Introdução

O levantamento da produção científica de um país permite estudar um dos aspectos que podem ser qualificados como resultados mensuráveis do amplo sistema de ciência, tecnologia e inovação. Acompanhar o fluxo de comunicação científica das diversas áreas facilita o processo de avaliação da pesquisa, cujas características são tão diversificadas quanto o é a própria ciência.

A análise quantitativa da ciência, que se dá a partir de indicadores, é tida como complementar à análise qualitativa, realizada pelos pares em diversas etapas, desde a formação, com a entrada e progressão na carreira de cientista, até a avaliação das pesquisas empreendidas, consubstanciada em manuscritos e títulos.

Nesse sentido, diferentes abordagens foram consideradas para avaliar e analisar o impacto da indexação de revistas científicas em contextos regionais ou locais. Testa (2011) destaca a expansão da cobertura de revistas nacionais na *Web of Science* (WoS) de diversas regiões. O fenômeno foi analisado em diversos trabalhos e contextos, como, por exemplo: Quevedo-Blasco (2013) apresenta uma visão geral das revistas ibero-americanas de Psicologia indexadas no *Journal Citations Reports* (JCR) de 2011; Purnel e Quevedo-Blasco (2013) analisam o impacto das revistas espanholas na WoS, identificando após 2005 um expressivo incremento, tanto da produção científica quanto da indexação espanhola; Collazo-Reyes (2013) apresenta e discute o aumento do número de revistas indexadas na WoS, no conjunto de revistas da América Latina e do Caribe, devido principalmente à política de abertura dos critérios de seleção da WoS.

Similarmente, nesses últimos anos, estudos bibliométricos foram realizados para analisar a produção científica indexada na WoS de diferentes áreas, em diversos países: Canadá (Wolfram, 2012), Colômbia (Rojas-Sola & de San-Antonio-Gomez, 2010), Croácia (Pikic *et al.*, 2012), Coreia (Huh, 2013; Fink *et al.*, 2014), Malásia (Abrizah *et al.*, 2013) e Portugal (Costa *et al.*, 2013).

No Brasil, a análise de indicadores da ciência remonta aos anos 70 (Morel, R. & Morel, C., 1977), quando o país apresentava uma produção na casa das centenas de artigos indexados nas bases do *Institute for Scientific Information* (ISI), atual *Thomson Reuters*, e ocupava posição próxima à 30ª no *ranking* mundial.

Leta (2011) apresenta um conjunto de 16 estudos que têm como foco indicadores de resultado da ciência brasileira, publicados em revistas, e cujo período de análise cobre pelo menos 5 anos. Um deles é o já mencionado estudo de Morel e Morel, enquanto três outros trabalhos cobrem a década de 1990, e outros doze abrangem o período de 2000 em diante, o que aponta a atualidade desses estudos. A autora afirma ainda que as agências de fomento, como o Ministério de Ciência e Tecnologia (MCT) e a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) têm se encarregado do esforço de divulgação de estatísticas e indicadores de ciência e tecnologia, após a virada do milênio – porém, iniciativas de análise do fenômeno, como visto, são esporádicas.

Durante a última década pôde-se notar maior envolvimento da comunidade científica nacional com o tema dos indicadores bibliométricos, o que decorreu de implementações dos critérios de avaliação do amplo sistema orquestrado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Criado em 1976 para avaliação de desempenho de programas de pós-graduação (Guimarães & Humann, 1995), vem tendo seus critérios aprimorados gradativamente, passando a apresentar uma abordagem mais objetiva a partir de 1998 (Souza & Paula, 2002).

Analisando os critérios de avaliação das diversas áreas no triênio 2007-2009, Miranda e Mugnaini (2013) observaram que a produção intelectual é um dos cinco itens que compõem o formulário. Algumas áreas atribuem pontuação de 40%, e outras 35% - os demais itens medidos são a proposta do programa, o quadro docente,

as dissertações e teses de alunos e a inclusão social. Já o peso da publicação em revistas científicas varia de 30% a 65%, peso esse que, multiplicado pelo peso da produção intelectual, resultará no peso da publicação em periódicos na avaliação total do programa, que varia de 12% (áreas de Artes e Música) a 26% (Ciência da Computação e Matemática, Probabilidade e Estatística). Segundo os autores, um critério predominante tem sido o fator de impacto do JCR, que crescentemente vem servindo como o principal critério de classificação das revistas consideradas mais importantes, seguido da indexação das revistas em índices de citação como WoS, Scopus e *Scientific Electronic Library Online* (SciELO) (Mugnaini & Sales, 2011).

Essa conjuntura que se delineou no Brasil expressa a relação direta entre indicadores bibliométricos e política científica, tornando os estudos quantitativos da ciência uma utilidade nacional, porém mantendo excessiva dependência de produtos e sistemas de indicadores bibliométricos oferecidos por empresas como *Thomson* e *Elsevier* (Mugnaini, 2013).

Em contrapartida, as revistas nacionais, que desde o início da década de 90 recebiam financiamento das agências de fomento, visando estimular a disseminação do conhecimento gerado por teses, assim como a produção científica de áreas de interesse local (Guimarães & Humann, 1995), têm logrado significativo reconhecimento nacional e internacional, sinalizando uma mudança de papel no fluxo de comunicação científica nacional. Nesse sentido, desde 1998, o Projeto SciELO, atuante junto aos editores nacionais, tem conferido não apenas maior visibilidade à pesquisa nacional (para a própria nação), mas dinamizado a editoria de revistas científicas (Packer & Meneghini, 2007). Também, a partir de 2005, as revistas vêm demonstrando indexação progressiva na *Web of Science* (Testa, 2011), proporcionando o aumento da produção brasileira naquela base (Leta, 2011). Além disso, desde o início de 2014, passam a compor o SciELO *Citation Index* as citações recebidas de revistas indexadas na *Web of Science* (Packer, 2014), agregadas às citações nacionais (já computadas pelo Projeto SciELO desde o ano 2000), passando a unir as bases regionais à denominada Principal Coleção da *Web of Science* (Thomson Reuters, 2014).

O processo de seleção de revistas do SciELO se baseia em critérios de mérito científico rigorosos, entre

outros aspectos, resultando que houve casos de revistas que conseguiram indexação antes na *Web of Science*. Mugnaini e Sales (2011) observaram que a validade de tais critérios tem sido amplamente reconhecida pela comunidade científica nacional, podendo ser notada pelo fato de 37% (17) das áreas de avaliação terem considerado a indexação no SciELO como um critério para classificação de revistas de mais alto estrato (A1).

Packer (2011) compara a posição dos vinte países mais bem posicionados no *ranking* mundial, tanto em número de artigos quanto em número de revistas indexadas na *Web of Science*, no ano de 2010. O autor apresenta interessantes constatações a respeito da diferença de posição que um mesmo país ocupa em cada *ranking*. Como exemplo destacam-se Holanda e Suíça, cuja posição no *ranking* de revistas é cerca de dez posições à frente do *ranking* de artigos, devido à concentração de títulos de editoras comerciais desses países. O autor apresenta como caso inverso a situação da Coreia do Sul, Índia e China, com posição de destaque no *ranking* de artigos, o que, segundo ele, pode decorrer de inconsistência na qualidade da editoração científica nacional ou na aplicação dos critérios exigidos para indexação. Sobre o Brasil, cuja posição no *ranking* de artigos (13^º) está pouco melhor que no *ranking* de revistas (15^º), Packer conjectura ser devido à subindexação das revistas nacionais.

Por outro lado, Leta (2012) esclarece que a maior presença dos autores brasileiros na *Web of Science* - em grande parte, fruto da presença de novos títulos nacionais na base - ocasionou a diminuição do percentual de artigos brasileiros em inglês, de 93,6% (2001) para 82,4% (2010). Diante dessa constatação, alerta para a necessidade de iniciativas complementares, como a própria internacionalização das revistas - que poderiam passar a publicar artigos na língua franca - e a proficiência de pesquisadores, para que a ciência ganhe audiência internacional.

Considerando o contexto delineado neste estudo, de acordo com os tópicos relacionados que compuseram o escopo temático das cinco últimas conferências da (Internacional Society for Scientometrics and Informetrics, 2014, *online*), observa-se que tanto "avaliação da política científica" quanto "indicadores de ciência e tecnologia" são tópicos gerais presentes em cada edição: Suécia (2005), Espanha (2007), Brasil (2009), África do Sul (2011) e Áustria (2013). Já os "estudos de produtividade em

nível nacional” deixaram de figurar na última edição do evento; por outro lado, é na edição brasileira que o tópico sobre “comparação entre bases de dados e cobertura” surge, perdurando até 2013; ademais, apenas a edição brasileira considerou o tópico “revistas não-anglófonas”, denotando dedicação exclusiva ao tema.

Tais constatações permitem afirmar que estudos de produção científica nacional, com vistas à geração de indicadores para subsídio à Política Científica, são alvo de interesse mundial, tendo a discussão sobre cobertura de bases de dados uma crescente atenção, (o que pode se dever à recente alternativa oferecida pela base Scopus, em relação à hegemonia da *Web of Science*). Contudo o tema relacionado às revistas não-anglófonas (e sua internacionalização) encontra menor interlocução, denotando assim a necessidade de ser discutido em âmbito nacional.

Diante do exposto, este estudo pretende analisar a produção científica nacional, a partir da agregação da produção científica brasileira registrada nas bases WoS e SciELO, considerando a influência de três fenômenos concomitantes observados por Mugnaini *et al.* (2014): (1) o crescimento do número de publicações em revistas internacionais, como resultado da política científica vigente há mais de uma década; (2) a crescente indexação de revistas brasileiras na WoS, na última década; e (3) o aumento da coleção de revistas exclusivamente indexadas na base SciELO.

Além de uma análise de produtividade, este estudo investiga o aumento, fluxo e dispersão da produção científica nacional. Para tanto, define as Zonas de Bradford das diversas áreas temáticas, buscando avaliar o aumento do número de revistas em cada uma das zonas, assim como a proporção de artigos publicados em revistas nacionais, ao longo de cinco triênios de produção. A abordagem busca oferecer uma caracterização da importância das revistas de diferentes zonas, na medida em que concentram maior ou menor percentual de artigos de determinada área, oferecendo assim novos parâmetros para a avaliação da produção científica nacional.

Métodos

O período de análise foi delimitado a cinco triênios completos de avaliação de programas de Pós-

-Graduação brasileira (1998 e 2012). O início do período coincide tanto com a criação da SciELO quanto com o momento em que a Capes passa a dar à avaliação um caráter mais objetivo, enquanto o fim do período coincide com o final do último triênio. Esse período de tempo é abrangente o suficiente para que possam ser observados possíveis efeitos, decorrentes da adoção de critérios de avaliação específicos, no perfil da produção científica nacional, a cada triênio, sendo esse o motivo das análises haverem sido realizadas trienalmente.

As fontes de dados foram as bases WoS e SciELO, consistindo dos registros bibliográficos de artigos (originais ou de revisão) que apresentassem pelo menos um autor com endereço institucional do Brasil, considerando variações que atendessem a diversos idiomas (BRAZIL, BRASIL ou BRESIL).

Da WoS foi recuperado um total de 402 845 registros, dos quais selecionaram-se somente os artigos originais ou de revisão publicados em revistas, restando 310 195 deles, uma vez descartadas as publicações em anais de conferências. A recuperação foi realizada automaticamente em setembro de 2013, com o auxílio de um macro de programação desenvolvido por Milanez (2011). Já os registros bibliográficos dos artigos originais ou de revisão da SciELO foram entregues pelos mantenedores da base, mediante solicitação, em janeiro de 2014, somando 162 582. Descartadas as sobreposições entre as bases, chegou-se a um total de 396 650 registros.

As revistas exclusivas de cada base foram identificadas como tal (“SciELO” ou “WoS”), e revistas indexadas em ambas receberam identificação “SciELO/WoS”, sendo os dados destas últimas obtidos da base que apresentasse o maior número de artigos. Essa diferença pode advir do fato de a revista haver sido indexada em uma base alguns anos antes da indexação na outra.

Foi considerado como ano de indexação das revistas SciELO aquele que a própria SciELO divulga como ano de entrada. Tal procedimento fez com que revistas que tiveram fascículos anteriores inseridos (indexação retrospectiva) fossem consideradas no início do período da análise como não indexadas (em 1998 somaram 1 500 artigos, chegando a 2 180 em 2001, e decrescendo até o final do período, chegando a zero em 2012). Por outro lado, revistas indexadas na SciELO em 2013, cujos artigos de 2012 foram indexados retrospectivamente, não

compuseram o *corpus* da pesquisa - a não ser no caso de a publicação ter sido indexada até 2012 na WoS. No caso da WoS, por falta de acesso à informação sobre o ano exato de indexação, foi considerado o ano mais antigo, segundo os artigos presentes na base.

A organização da produção segundo áreas do conhecimento consistiu na adoção da classificação do *Essential Science Indicators*® (ESI), que agrupa aproximadamente 250 áreas de classificação de revistas da WoS (ambas da *Thomson Reuters*) em 22 áreas - às quais foi adicionada a área de Ciências Humanas, originalmente não contemplada naquele produto. No caso das revistas indexadas na WoS, foi utilizada a lista de revistas (*Journal List*) que a própria instituição disponibiliza em livre acesso no *Science Watch* (Thomson Reuters, 2013). Já para as revistas da SciELO, utilizou-se a classificação segundo categorias da WoS, para posterior reclassificação segundo áreas do ESI, consultando as definições temáticas (*Field Definitions*) do *Science Watch*.

Para análise da comunicação científica de cada área, com vistas à determinação do conjunto de revistas importantes para as áreas, foi utilizada a Lei de Dispersão de Bradford (Bradford, 1961). O uso dessa metodologia permitiu identificar três conjuntos de revistas (ou Zonas de Bradford) de acordo com a quantidade de artigos publicados, sendo que cada zona concentra um terço da produção de uma área específica. A Zona 1 reúne um número reduzido de revistas, com grande quantidade de artigos cada uma; a Zona 2 apresenta um número maior de revistas, porém com menores quantidades de artigos cada uma; e a Zona 3 mostra maior dispersão, com muitas revistas, cada qual com poucos artigos. A primeira zona contém as revistas mais representativas do processo de comunicação científica das áreas; a segunda zona, as revistas em consolidação; e a terceira, as revistas novas, as que estão deixando de ser utilizadas e as que são utilizadas ocasionalmente. Cada revista é classificada em uma única área pela ESI. Definidas as Zonas de Bradford da produção científica das diferentes áreas em cada triênio, foi possível analisar as variações do número de revistas de cada zona, bem como da proporção de artigos em revistas nacionais.

A nacionalidade da revista utilizou a seguinte agregação, baseada na região de publicação declarada pelo editor: nacional (Brasil), regional (América Latina &

Caribe, exceto Brasil) e internacional (demais países não considerados nas categorias anteriores).

Resultados e Discussão

O fluxo de comunicação científica no Brasil apresenta um cenário diversificado no período analisado, considerando a indexação das revistas brasileiras nas bases SciELO e WoS (Mugnaini *et al.*, 2014). Tal quadro revela a entrada crescente de revistas indexadas exclusivamente na SciELO em todos os triênios, com destaque para o quarto triênio, quando um expressivo número de revistas brasileiras foi incluído na WoS, com a expansão realizada pela Thomson Reuters (Testa, 2011), entre 2006 e 2008.

Contudo, considerando-se o fluxo de artigos nessas revistas, nota-se que a maior parte da produção nacional se dá em revistas indexadas somente na WoS (Mugnaini *et al.*, 2014): no primeiro triênio, período de criação da base SciELO, os artigos em revistas WoS representaram 73,1%, percentual decrescente que atinge o valor mínimo de 54,7% em 2007, e cresce ao longo do último triênio, atingindo 59,0% em 2012.

Certamente a maior presença de revistas brasileiras implicou aumento do número de artigos, e consequentemente, maior representatividade da pesquisa nacional na WoS. Por outro lado, é necessário analisar o aumento da produção brasileira, separando o aumento do número de artigos em revistas nacionais, do aumento em revistas internacionais.

Na Tabela 1 observa-se que a taxa de crescimento percentual da produção em revistas estrangeiras indexadas na WoS apresenta crescimento estável entre os triênios 2 e 4, com desaceleração no último triênio (com mais de 66 mil artigos publicados), sendo a soma de crescimento no período de 124,3%. A produção de revistas internacionais na base SciELO ou revistas SciELO/WoS - que apresentam significativo aumento - diz respeito às coleções dos projetos SciELO de outros países, como África do Sul, Espanha, Portugal e algumas revistas do SciELO Saúde Pública.

É entre as revistas brasileiras (categoria nacional) indexadas na SciELO ocorre a maior taxa de crescimento. Aquelas indexadas também na WoS somam um cresci-

mento de 442,6% no período, com destaque para os triênios 2 e 4, cujo aumento decorre diretamente da indexação das revistas na WoS, e apenas 16,2% no último triênio (com mais que 29 mil artigos). Quanto às revistas indexadas somente na SciELO, a maior taxa de crescimento se dá do triênio 2 (quando da inclusão do primeiro conjunto de revistas da base), decrescendo até uma taxa negativa de -18,1% no triênio 4 (quando diversas revistas entram também na WoS), voltando a crescer (23,6%) no último triênio (com a publicação de mais de 19 mil artigos) e somando 345,9% de crescimento no período. A produção em revistas nacionais indexadas exclusivamente na WoS mostra aumentos altíssimos, claramente resultante da entrada das revistas naquela base nos triênios 3 e 4.

Já as revistas regionais, que correspondem àquelas publicadas em outros países latino-americanos,

apresentam menor quantidade de artigos, também demonstrando aumentos altíssimos quando elas são indexadas também na WoS. Contudo, entre as revistas exclusivas da SciELO, nota-se um aumento constante, acumulando 402,9% no período completo, denotando a importância das revistas da região para a ciência brasileira - ainda que tenham pequena expressão em números absolutos.

Dada a menor quantidade de revistas SciELO, o aumento da base ou indexação na WoS apresenta sensível taxa de crescimento. Esse fenômeno é acentuado pelo alto percentual de artigos de brasileiros nas revistas nacionais como um todo. Segundo Collazo-Reyes (2013), no período entre 2005 e 2011 as revistas SciELO/WoS brasileiras apresentaram 88,8% de endogenia - ou seja, apenas 10,2% artigos de autores estrangeiros -, realidade que contrasta com países como Argentina,

Tabela 1. Distribuição e crescimento percentual trienal da produção científica brasileira, segundo a região de publicação da revista e a base de indexação - período de 1998-2012.

Base	Região de publicação da revista						
	Internacional		Regional		Nacional		Total
	Frequência	Crescimento %	Frequência	Crescimento %	Frequência	Crescimento %	
Triênio							Frequência
WoS	212 593	124,3	1 135	1 155,7	12 665	1644,1	226 393
1	22 522	-	160	-	3 201	-	25 883
2	30 171	34,0	67	-58,1	45	-98,6	30 283
3	40 319	33,6	32	-52,2	560	1144,4	40 911
4	53 206	32,0	437	1 265,6	3 683	557,7	57 326
5	66 375	24,8	439	0,5	5 176	40,5	71 990
SciELO/WoS	8 873	175,1	1 881	14 027,1	70 434	442,6%	81 188
1	731	-	1	-	1 986	-	2 718
2	1 047	43,2	138	13700,0	5 680	186,0	6 865
3	1 807	72,6	191	38,4	8 710	53,3	10 708
4	2 255	24,8	703	268,1	25 003	187,1	27 961
5	3 033	34,5	848	20,6	29 055	16,2	32 936
SciELO	1 478	310,9	2 702	402,9	61 316	345,9	65 496
1	81	-	85	-	2 746	-	2 912
2	106	30,9	159	87,1	10 289	274,7	10 554
3	180	69,8	377	137,1	17 049	65,7	17 606
4	355	97,2	730	93,6	13 969	-18,1	15 054
5	756	113,0	1 351	85,1	17 263	23,6	19 370
Não indexada	3 079		659		18 835		22 573
Total	226 023	57,1	6 377	1,6	163 250	41,3	395 650

Fonte: elaborado pelos autores (2014).

Nota: Região da revista, baseada no país de publicação, declarado pelo editor: nacional (Brasil), regional (América Latina & Caribe, exceto Brasil) ou internacional (demais países não considerados nas categorias anteriores); SciELO: *Scientific Electronic Library Online*; WOS: *Web of Science*.

Chile e México, cujo percentual de endogenia é próximo de 67%, segundo o autor.

Esses resultados reforçam a já citada necessidade de internacionalização das revistas científicas (Leta, 2011), uma vez que atualmente o ingresso na WoS não depende do idioma. Por outro lado, Collazo-Reyes (2013) destaca que a maior presença da pesquisa brasileira na WoS fez com que no período de 2005-2011 o português se tornasse o segundo idioma mais frequente (após o inglês) dos artigos publicados em revistas latino-americanas. O debate ora suscitado relembra os questionamentos da década de 1990, quando se discutia a exclusão de países não-anglófonos, porém agora exigindo o passo seguinte: internacionalização da ciência brasileira.

Nesse sentido, a análise da produção brasileira nas áreas do conhecimento pode ser mais detalhada, ao se comparar o percentual de artigos em revistas internacionais, assim como sua variação ao longo do tempo: observou-se pouca alteração, com crescimento médio de 57%, e leve crescimento durante o triênio 5, que compreende o período de 2010 a 2012 (Mugnaini *et al.*, 2014). Tal constatação sugere que a desaceleração do crescimento da base SciELO (Packer & Meneghini, 2007), assim como o aumento da produção em revistas internacionais crescentemente estimulado pela avaliação nacional, está influenciando o crescimento relativo da produção em revistas internacionais.

Torna-se então necessário considerar que a produção científica distribui-se em conjuntos de revistas, que diferem quanto à magnitude do número de artigos, devendo-se ainda distinguir a produção nas diversas áreas.

A Figura 1 apresenta a comparação do tamanho das Zonas de Bradford total e em cada área. Em ordem decrescente do número de revistas que compõem a zona 1, observa-se primeiramente que as 113 revistas que compõem são responsáveis por 1/3 da produção nacional no período entre 1998 e 2012, enquanto as zonas 2 e 3 apresentam, respectivamente, 593 e 8 383 revistas. As áreas de Medicina Clínica, Engenharia e Matemática são as que apresentam maior número de revistas na zona 1, com 26, 21 e 19 revistas respectivamente. Já nas áreas de Ciência do Espaço e Multidisciplinar, identifica-se apenas uma revista na zona 1 - revistas que publicaram no período completo um total de 898 e 744 artigos, respectivamente. Considerando a linha

de tendência que sobrepõe as barras, pode-se observar a área Multidisciplinar é a menos expressiva, com um total de 1 676 artigos, em contraste com Medicina Clínica (87 036), Ciência de Plantas e Animais (50 148), Ciências Sociais (31 332), Física (30 467), Química (29 514) e Ciências Agrárias (28 201).

Quanto à comparação entre o tamanho das zonas, por meio da Correlação de Spearman entre o percentual de revistas por zona em cada área, observa-se maior intensidade na correlação negativa entre as zonas 2 e 3 (-0,983), significando que as áreas em que a zona 3 é mais representativa são aquelas nas quais a representatividade da zona 2 é menor. Destacaram-se como áreas de maior dispersão, ou seja, com maior percentual de revistas na zona 3: Ciências Sociais (zona 2 com 2,6% do total de revistas, e zona 3 com 96,8%), Psiquiatria/Psicologia (zona 2 com 2,4%, e zona 3 com 96,4%) e Economia e Negócios (zona 2 com 3,0%, e zona 3 com 95,6%). Contrariamente, as áreas com maior zona 2 são: Imunologia (zona 2 com 14,3%, e zona 3 com 80,2%), Matemática (zona 2 com 14,6%, e zona 3 com 81,0%) e Ciência da Computação (zona 2 com 13,3%, e zona 3 com 82,2%).

Tendo conhecido a distribuição do número de revistas e de artigos segundo as bases de indexação e a magnitude das zonas das diversas áreas, é necessário analisar a presença das revistas nacionais na publicação da ciência brasileira.

Para tanto, considerou-se, na Figura 2, não a base de indexação, mas o país de publicação das revistas nacionais, avaliando-se o percentual dos artigos publicados nessas revistas, nas zonas de cada área. Uma primeira observação é que a zona 1 apresenta o maior percentual de artigos em revistas nacionais na maioria das áreas. A zona 2, por sua vez, apresenta maior percentual que a zona 3, com exceção apenas da área de Microbiologia. Esse comportamento decrescente é bem demarcado quando se observam os percentuais das zonas correspondentes à produção nacional total (Figura 1), representada como Brasil, que anteriormente encontrava-se na primeira posição).

Ainda na Figura 2, pode-se notar um conjunto de cinco áreas cujas zonas 1 e 2 são predominantemente de revistas nacionais: Ciências Humanas, Economia e Negócios, Psiquiatria/Psicologia, Ciências Agrárias e Ciências Sociais. Considerando que a zona 3 apresenta

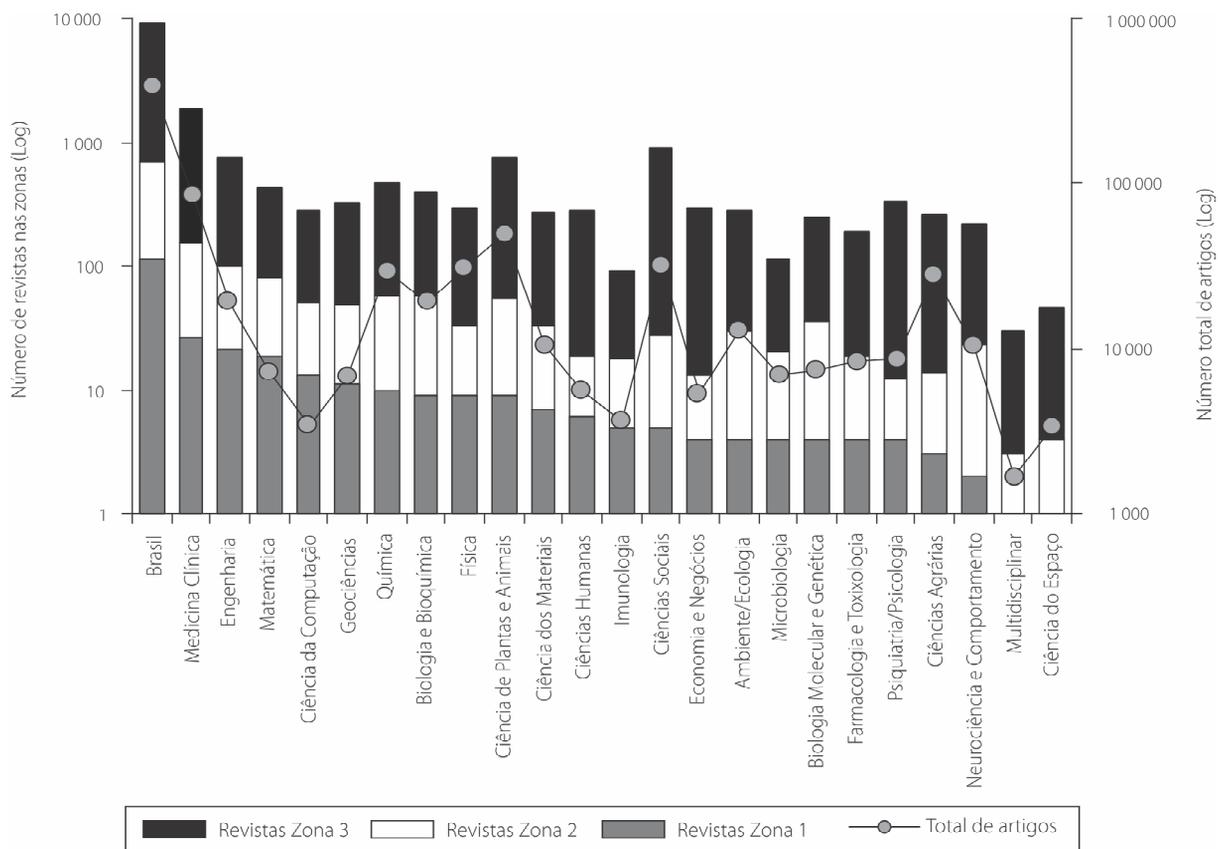


Figura 1. Produção científica brasileira, segundo o tamanho das Zonas de Bradford (número de revistas) e o número de artigos, por áreas temáticas - período de 1998-2012.

Fonte: Elaborado pelos autores (2014).

mais dispersão entre um conjunto maior de revistas, observa-se que, dessas áreas, apenas Ciências Humanas apresenta mais alto percentual (57,9%) de artigos em revistas nacionais, sendo a área de menor internacionalização na sua produção. A área de Ciências Sociais apresenta comportamento similar, diferindo apenas no fato de que, na zona 2, a produção em revistas nacionais representa 90,2%, e na zona 3, 53,6%. Entre as demais áreas pode-se observar que a zona 3 é predominantemente composta de artigos em revistas internacionais, concentrando a maioria das áreas de saúde e biológicas mais ao centro do gráfico, com a zona 1 apresentando percentual de artigos em revistas nacionais variando de 71,7% (Medicina Clínica) a 94,1% (Biologia e Bioquímica). Já no extremo direito predominam as áreas de exatas e da Terra, com zona 1 variando de 0% (Física, Imunologia e Ciência do Espaço) a 68,8% (Química). Finalmente, destaca-se o maior percentual de artigos em revistas

nacionais na zona 2, nas áreas de Ciência de Plantas e Animais e Farmacologia e Toxicologia.

A ordenação das áreas segundo o percentual de artigos brasileiros na zona 1, observada na Figura 2, vai ao encontro das evidências apresentadas por Leite *et al.* (2011), que analisaram o percentual de pesquisadores com perfil de publicação predominantemente internacional, com base em seus currículos Lattes. Contudo, o presente estudo permite análises mais criteriosas das revistas que compõem as zonas 2 e 3, podendo revelar peculiaridades do fluxo de comunicação científica das áreas, à medida que revistas internacionais começam a surgir e serem consolidadas em área de perfil predominantemente nacional e vice-versa.

Pelo fato de esta análise não pretender entrar no nível das revistas em si, mas da evolução das zonas, apresentam-se a seguir duas tabelas que, com base nos dados utilizados para a elaboração das duas figuras

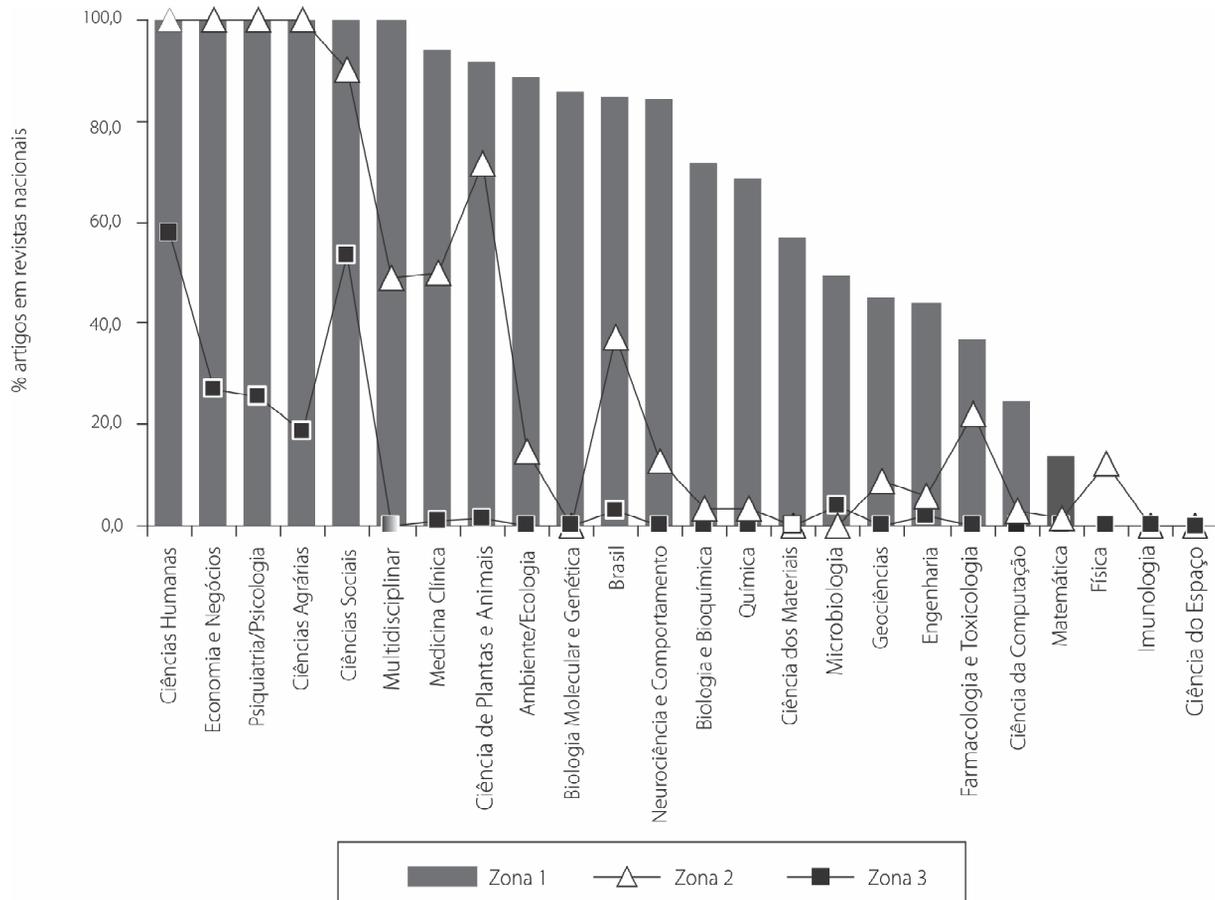


Figura 2. Porcentagem de artigos em revistas nacionais, por Zonas de Bradford e área - período de 1998-2012.

Fonte: Elaborado pelos autores (2014).

anteriores, permitem a análise da variação do tamanho das zonas de cada área, ao longo dos triênios.

Na Tabela 2 pode-se observar que, considerando o total da produção nacional (Brasil), há aumento no número de revistas das zonas, entre todos os triênios.

De maneira geral, como é de se esperar, a zona 3 é a que mais apresenta aumento (+) do número de revistas entre os triênios. Geralmente nessa zona agregam-se revistas novas, nas quais a comunidade de uma área passa a publicar. Com base nos números mínimo e máximo de revistas dos triênios, ao longo do período, observa-se que áreas como Ciências Humanas, Ciências Sociais e Multidisciplinar apresentam aumento destacado. Por outro lado, algumas áreas têm o número de revistas reduzido (-) entre alguns triênios, como exemplo a Física, cuja redução é observada entre os triênios 2-3 e 4-5.

À medida que se sobe para as zonas 2 e 1, respectivamente, observa-se menor número de aumentos (+) entre as áreas, bem como maior número de diminuições (-).

Os números mínimo e máximo das áreas na zona 2 permitem destacar as áreas com maior aumento, em ordem decrescente: Ciências Humanas, Ciências Agrárias, Economia e Negócios, Ciências Sociais e Psiquiatria/Psicologia. São exatamente as mesmas áreas que na Figura 2 apresentaram predominância de artigos em revistas nacionais nessa zona.

Observando a Tabela 3, a seguir, pode-se considerar que, destas áreas, Ciências Agrárias e Psiquiatria/Psicologia apresentam redução desse percentual no último triênio, enquanto Ciências Sociais apresenta oscilação.

A ordenação da Tabela 2 se baseia no número máximo de revistas por triênio, na zona 1. Observa-se que, na base, com menor número de revistas, encontra-se a maioria das áreas recém-mencionadas, mais a área de Ciência do Espaço e Multidisciplinar. Já as exatas e da terra estão no topo da Tabela, com exceção de Medicina Clínica e Ciência de Plantas e Animais - na verdade as áreas de saúde e biológicas encontram-se dispersas entre o conjunto.

Ainda sobre a Tabela 2, é importante mencionar que a maior ocorrência de aumento das áreas se dá entre os triênios 2-3 (na zona 3), 3-4 (na zona 2) e 4-5 (na zona

1). Pode-se associar esse fenômeno ao sistema de avaliação coordenado pela Capes desde 1998 (Souza & Paula, 2002), uma vez que seu efeito mais imediato se dá entre as revistas de menor frequência (das zonas 3 das áreas) entre os triênios 2-3, ou seja, entre os anos 2003 e 2004. Conseqüentemente, chega às zonas 2 na transição das avaliações trienais entre os anos 2006 e 2007, e, finalmente, às zonas 1 entre os anos 2009 e 2010. Trata-se do *Qualis*, que consiste na classificação dos veículos (revistas, anais de conferência e livros) utilizados pelos programas de pós-graduação para publicação de suas pesquisas (Coordenação de Aperfeiçoamento de Pessoal

Tabela 2. Variação (positiva, negativa ou nula), mínimo e máximo do número de revistas das zonas das diversas áreas entre os triênios - período de 1998-2012.

Áreas	Número de revistas																	
	Zona1				Zona2				Zona3									
	1-2	2-3	3-4	4-5	mín	máx	1-2	2-3	3-4	4-5	mín	máx	1-2	2-3	3-4	4-5	mín	máx
Brasil	+	+	+	+	56	122	+	+	+	+	357	564	+	+	+	+	3.375	6.233
Medicina Clínica	+	+	+	+	9	30	-	+	+	-	66	109	+	+	+	+	662	1.357
Engenharia	-	=	+	+	14	19	+	+	+	+	58	69	+	+	+	+	284	471
Matemática	+	+	=	=	13	17	+	+	+	-	42	57	+	+	+	+	148	274
Ciência da Computação	+	=	+	+	3	14	+	=	+	+	21	36	+	+	+	+	72	159
Química	+	-	-	+	8	11	+	-	+	-	35	42	+	+	+	+	221	301
Ciência de Plantas e Animais	+	+	+	+	4	11	-	-	+	+	22	45	+	+	+	+	330	556
Geociências	+	-	=	=	8	10	+	-	+	-	25	33	+	+	+	+	111	208
Biologia e Bioquímica	+	-	+	+	5	10	+	=	+	+	28	48	+	+	+	+	185	262
Física	-	=	+	-	7	8	=	-	+	+	18	22	+	-	+	-	168	197
Ciência dos Materiais	-	=	=	+	6	8	-	=	+	+	19	23	+	+	+	-	109	168
Ciências Humanas	=	+	+	+	2	8	+	+	+	+	3	12	-	+	+	+	44	147
Imunologia	=	+	+	+	4	7	=	+	+	+	9	13	+	+	-	+	33	53
Ambiente/Ecologia	-	=	=	=	3	7	=	=	+	+	19	27	+	+	+	+	90	205
Ciências Sociais	=	+	+	+	3	7	+	+	+	+	9	23	+	+	+	+	173	548
Neurociência e Comportamento	=	+	+	+	1	6	+	+	+	+	11	23	+	+	+	+	104	154
Microbiologia	-	+	-	+	3	5	+	+	+	+	10	17	+	+	+	+	46	79
Psiquiatria/Psicologia	=	+	+	=	2	5	+	+	+	+	4	10	+	+	+	+	98	208
Biologia Molecular e Genética	+	-	-	+	1	5	+	-	=	=	14	26	+	+	+	+	88	160
Ciências Agrárias	+	=	+	+	1	5	=	+	+	+	4	12	+	+	+	-	101	183
Economia e Negócios	+	+	+	+	1	5	=	+	+	+	3	9	+	+	+	+	60	182
Farmacologia e Toxicologia	=	=	=	+	3	4	+	+	+	+	9	18	+	+	+	+	63	136
Ciência do Espaço	-	=	=	+	1	2	+	=	+	-	2	4	=	+	-	+	19	32
Multidisciplinar	=	=	=	=	1	1	=	=	=	=	2	2	-	+	+	+	6	23

Fonte: Elaborado pelos autores (2014).

Nota: Foram utilizados sinais indicando se as mudanças entre triênios subsequentes foram de "+" (crescimento), "-" (decréscimo) ou "=" (inalterado). Valores mínimo (mín) e máximo (máx) correspondem aos valores extremos observados entre os triênios, não importando qual.

de Nível Superior, 2014). Em relação às revistas, consiste em uma lista disponibilizada às comunidades científicas de quase 50 áreas de avaliação, com as seguintes classificações: A1 (mais alta) e A2, B1 a B5 (mais baixa) e C (não científicas). A lista de cada área aumenta a cada triênio, à medida que a produção se dispersa por novas revistas, conforme observado na zona 3. Além disso, a classificação segundo estratos distingue revistas, ação que pode explicar o aumento do número de artigos em algumas delas, repercutindo assim nas zonas 1 e 2.

Na Tabela 3, observando o total da produção nacional (Brasil), nota-se alta concentração de artigos em

revistas nacionais na zona 1 (mínimo 79% e máximo 86%), valores que decrescem drasticamente para a zona 2 (mínimo 24% e máximo 43%) e se reduzem a 2% (mínimo e máximo) para a zona 3.

Este é outro efeito relacionados com os critérios de avaliação, uma vez que o estímulo à publicação em revistas internacionais vem crescentemente se alastrando ao longo do período, e no quarto triênio (2007-2009) o Fator de Impacto passa a ser adotado por diversas áreas, como Ciências Sociais (Administração, Contabilidade & Turismo), dobrando o valor mínimo do indicador de 0,5 para 1,0 no último triênio. Ressalta-se que, no presente

Tabela 3. Variação (positiva, negativa ou nula), mínimo e máximo do percentual de artigos em revistas nacionais nas zonas das diversas áreas, entre os triênios - período de 1998-2012.

Áreas	Percentual de artigos em revistas nacionais																	
	Zona1						Zona2						Zona3					
	1-2	2-3	3-4	4-5	mín	máx	1-2	2-3	3-4	4-5	mín	máx	1-2	2-3	3-4	4-5	mín	máx
Ciências Agrárias	=	=	=	=	100%	100%	=	=	=	-	86%	100%	-	-	+	-	17%	43%
Ciências Humanas	=	=	=	=	100%	100%	=	=	=	=	100%	100%	+	-	+	+	31%	66%
Economia e Negócios	=	=	=	=	100%	100%	+	=	=	=	91%	100%	+	+	-	-	0%	45%
Multidisciplinar	=	=	=	=	100%	100%	=	=	+	+	0%	69%	=	=	=	=	0%	0%
Psiquiatria/Psicologia	=	=	=	=	100%	100%	=	=	=	-	95%	100%	+	-	-	-	16%	43%
Ciências Sociais	=	=	=	-	89%	100%	+	-	+	-	88%	93%	+	-	-	-	49%	63%
Medicina Clínica	=	=	=	=	88%	100%	+	-	-	+	35%	58%	-	+	+	-	0%	1%
Ciência de Plantas e Animais	=	=	=	=	88%	100%	+	-	+	-	36%	83%	+	+	-	-	1%	3%
Neurociência e Comportamento	=	-	-	-	65%	100%	=	+	+	-	0%	16%	=	=	=	=	0%	0%
Biologia Molecular e Genética	-	+	+	=	52%	100%	=	=	=	=	0%	0%	=	=	=	=	0%	0%
Biologia e Bioquímica	-	+	-	-	63%	89%	+	-	=	=	0%	7%	+	-	=	=	0%	1%
Ambiente/Ecologia	+	-	+	=	29%	100%	+	+	-	+	0%	28%	=	=	+	-	0%	2%
Brasil	+	+	+	-	79%	86%	+	+	+	-	24%	43%	+	-	+	-	2%	2%
Geociências	-	+	+	+	17%	76%	+	-	+	-	0%	6%	=	=	=	=	0%	0%
Química	-	+	+	-	64%	74%	-	+	+	-	0%	8%	=	=	=	=	0%	0%
Farmacologia e Toxicologia	+	+	-	+	0%	68%	=	=	+	-	0%	25%	-	=	=	=	0%	5%
Microbiologia	+	-	+	-	42%	66%	=	=	=	=	0%	0%	=	=	=	=	0%	0%
Ciência dos Materiais	+	-	+	-	40%	60%	=	+	-	=	0%	7%	=	=	=	=	0%	0%
Engenharia	+	-	-	+	25%	54%	=	=	+	+	0%	6%	=	=	=	+	0%	1%
Ciência da Computação	-	+	+	-	22%	31%	=	=	=	=	0%	0%	=	=	=	=	0%	0%
Matemática	=	+	-	+	0%	17%	+	-	+	-	0%	7%	=	=	+	-	0%	0%
Física	=	=	=	+	0%	9%	+	+	+	-	0%	15%	-	=	=	=	0%	1%
Ciência do Espaço	=	=	=	=	0%	0%	=	=	=	=	0%	0%	=	=	=	=	0%	0%
Imunologia	=	=	=	=	0%	0%	=	=	=	=	0%	0%	=	=	=	=	0%	0%

Fonte: Elaborado pelos autores (2014).

Nota: Foram utilizados sinais indicando se as mudanças entre triênios subsequentes foram de "+" (crescimento), "-" (decremento) ou "=" (inalterado). Valores mínimo (mín) e máximo (máx) correspondem aos valores extremos observados entre os triênios, não importando qual.

estudo, as Ciências Sociais incluem a área de Saúde Pública, entre cujos critérios de avaliação figura um considerável Fator de Impacto de valor 4,0 para classificação de revistas A1, o que pode explicar a diminuição percentual de publicações em revistas nacionais da zona 3, nos últimos quatro triênios. O mesmo se dá na área de Humanas, na qual se identifica o critério vinculado ao Fator de Impacto mínimo de 0,5 na área de Geografia, para classificação de revistas A1 (Mugnaini & Sales, 2011).

Analisando a Tabela como um todo, pode-se perceber que o número de aumentos (+) e diminuições (-) das diversas áreas, entre os triênios, é praticamente o mesmo, considerando-se a zona 1. Porém, ao se analisar a zona 2, os aumentos prevalecem, enquanto na zona 3, contrariamente, predominam os decréscimos. Tais constatações reforçam o efeito acima citado sobre a avaliação nacional.

Quando se observam as áreas específicas, ordenadas pelos percentuais máximo e mínimo de artigos em revistas nacionais da zona 1, nota-se as já mencionadas áreas nas quais há predomínio absoluto, assim como as áreas com maior percentual observado na zona 3. Dentre essas áreas, convém destacar as Ciências Agrárias, nas quais Fink *et al.* (2014) observaram um aumento de produção em relação à média mundial no período entre 2000 e 2009, o que segundo os autores não se expressou de maneira similar nas citações (apesar de serem o dobro da média mundial). Eles arriscam afirmar que talvez a ausência de aumento em citações pode se dever à especificidade da pesquisa nessa área, que pode estar direcionada a estudos de problemas locais, de menor interesse para o restante do mundo. Tal afirmação encontra respaldo neste estudo, na medida em que se observa a importância das revistas nacionais, predominando as zonas 1 e 2.

Em seguida vêm as áreas de saúde e biológicas que, em sua maioria, mostram diminuição desse percentual na zona 1, sinalizando um predomínio de internacionalização da pesquisa nos últimos triênios. Para as áreas de Biologia e Bioquímica, Biologia Molecular e Genética e Neurociência e Comportamento esses efeitos são observados em triênios anteriores, diferente de Medicina Clínica, cujo efeito só é notado entre os últimos dois triênios. Fink *et al.* (2014) identificam Ambiente/Ecologia, Farmacologia e Toxicologia e Microbiologia como

áreas que mantêm suas posições em relação à média mundial, fazendo-o de maneira balanceada, considerando produção e citação. Essas são áreas cujas zonas 2 e 3 são predominantemente internacionais, sendo que as duas primeiras vêm apresentando maior participação das revistas nacionais na zona 2 ao longo dos triênios.

Entre as áreas de exatas e da terra, que se agrupam na parte inferior da Tabela, destacam-se Geociências, Engenharia, Matemática e Física, por aumentos apresentados nas zonas 1 e 2, em diversos triênios. Na base da Tabela, Imunologia e Ciência do Espaço destacam-se pela publicação exclusivamente internacional.

Conclusão

Nos últimos quinze anos, a produção científica brasileira tem logrado melhores posições no cenário mundial, graças aos esforços que o país vem empreendendo em nível nacional. Os investimentos em formação e aperfeiçoamento de pesquisadores têm estimulado tanto a atividade científica em si quanto as atividades relacionadas à editoria de revistas nacionais. Os resultados deste estudo evidenciam que a produção tem avançado crescentemente às revistas estrangeiras, ao mesmo tempo que as revistas nacionais desempenham importante papel na infraestrutura de comunicação científica, tanto no âmbito nacional quanto internacional.

A complexidade associada ao processo de avaliação da produção científica se dá, em grande parte, pelas características das diversas áreas da ciência. Como se pôde perceber, o papel das revistas nacionais reforça ainda mais tais diferenças, na medida em que algumas áreas são menos internacionalizadas que outras. Há que se considerar ainda que este estudo limitou-se apenas aos artigos em periódicos, desconsiderando que algumas áreas costumam publicar mais em livros ou anais de eventos.

Outro aspecto diz respeito às zonas de Bradford que, num primeiro momento, revelaram que algumas áreas publicam num conjunto de revistas consideravelmente maior do que as demais, como Medicina Clínica. Outras áreas destacam-se pelo fato de apresentarem dispersão de sua produção em maior número de revistas da zona 3, que apresentam baixa frequência de artigos,

evidenciando a maior variedade de revistas para publicação, como é o caso de Ciências Sociais, Psiquiatria/Psicologia e Economia e Negócios.

Notou-se ainda um aumento mais generalizado do número de revistas das áreas, sucessivamente, na zona 3 (entre os triênios 2 e 3), zona 2 (entre os triênios 3 e 4) e zona 1 (entre os triênios 4 e 5). Associou-se a esse fenômeno o processo de avaliação Qualis, cuja lista de revistas aumenta a cada triênio, à medida em que a produção das áreas se dispersa por novos títulos, sendo esse o efeito mensurado no presente estudo. Além disso, a classificação do Qualis segundo estratos acaba distinguindo um grupo de revistas mais importantes, podendo explicar o aumento do número de revistas das zonas 1 e 2, com maior frequência de artigos.

Finalmente, a análise da variação do percentual da produção das áreas em revistas nacionais diferenciou três grupos de áreas. Na área de ciências sociais e humanidades, as zonas 1 e 2 são exclusivamente compostas por revistas nacionais, enquanto a zona 3 vem apresentando diminuição de percentual, devido ao esforço de publicação em revistas estrangeiras. Na área de Física, Ciência do Espaço e Imunologia, a produção se dá em revistas internacionais, não importando a zona. Já as demais áreas de ciências exatas e da terra, saúde e biológicas vêm apresentando aumento do percentual de

publicação em revistas nacionais nas zonas 1 e 2, permanecendo a zona 3 exclusivamente internacional.

Pôde-se concluir que o fluxo da produção tem sido influenciado pelas revistas nacionais, cujo reconhecimento tem sido atestado não apenas pela inserção na *Web of Science*, mas pela própria comunidade científica quando da definição dos critérios de classificação de revistas no Qualis.

A metodologia utilizada neste trabalho permite a análise da produção científica nacional, utilizando Zonas de Bradford sobre um longo período de tempo. Destaca-se que, em continuação, a pesquisa aplicará a mesma metodologia para avaliação do consumo de informação das publicações nacionais, com base nas referências bibliográficas, com o objetivo de prover a política científica nacional (Qualis) de indicadores dinâmicos e contextualizados das revistas utilizadas.

Agradecimentos

À Fundação de Amparo à Pesquisa do Estado de São Paulo (Projeto Jovens Pesquisadores 2012/00255-6) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (Projeto Universal 306046/2013-0 e Bolsa de Produtividade em Pesquisa 477246/2013-3).

Referências

Abrizah, A., *et al.* Citation performance of Malaysian scholarly journals in the Web of Science, 2006–2010. *Serials Review*, v.39, n.1, p.47-55, 2013.

Bradford, S.C. *Documentação*. Rio de Janeiro: Fundo de Cultura, 1961. p.196-216.

Collazo-Reyes, F. Growth of the number of indexed journals of Latin America and the Caribbean: The effect on the impact of each country. *Scientometrics*, v.89, n.1, p.197-209, 2013.

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. *Webqualis Capes*. 2014. Disponível em <<http://www.capes.gov.br/avaliacao/qualis>>. Acesso em: 12 jan. 2014.

Costa, T.; Vaz, F.; Lopes, C. The impact of the online knowledge library: Its use and impact on the production of the Portuguese academic and scientific community (2000-2010). In: Qualitative and Quantitative Methods in Libraries International Conference, 5., 2013, Roma. *Electronic proceedings...* Roma, 2013. Available from: <<http://repositorio.ispa.pt/handle/10400.12/2592>>. Cited: Aug. 8, 2014.

Fink, D. *et al.* S&T knowledge production from 2000 to 2009 in two periphery countries: Brazil and South Korea. *Scientometrics*, v.99, n.1, p.37-54, 2014.

Guimarães, J.A.; Humann, M. C. Training of human-resources in science and technology in Brazil: The importance of a vigorous postgraduate program and its impact on the development of the country. *Scientometrics*, v.34, n.1, p.101-119, 1995.

Huh, S. Citation analysis of the Korean Journal of Urology from Web of Science, Scopus, Korean Medical Citation Index, KoreaMed Synapse, and Google Scholar. *Korean Journal of Urology*, v.54, n.4, p.220-228, 2013.

International Society for Scientometrics and Informetrics. *Past Conferences*. 2014. Available from: <<http://www.issi-society.org/past.html>>. Cited: Jan. 12, 2014.

Leite, P.; Mugnaime, R.; Leta, J. A new indicator for international visibility: Exploring Brazilian scientific community. *Scientometrics*, v.88, n.1, p.311-319, 2011.

- Leta, J. Indicadores de desempenho, ciência brasileira e a cobertura das bases informacionais. *Revista USP*, n.89, p.62-67, 2011.
- Leta, J. Brazilian growth in the mainstream science: The role of human resources and national journals. *Journal of Scientometric Research*, v.1, n. 1, p.44-52, 2012.
- Milanez, D. H. *Nanotecnologia: indicadores tecnológicos sobre os avanços em materiais a partir da análise de documentos de patentes*. 2011. Dissertação (Mestrado em Ciência e Engenharia de Materiais) - Centro de Ciências Exatas e de Tecnologia de Materiais, Universidade Federal de São Carlos, 2011.
- Miranda, E.C.; Mugnaini, R. Scientific policy in Brazil: Exploratory analysis of assessment criteria. In: International Society of Scientometrics and Informetrics Conference, 14., 2013, Viena. *Proceedings...* Viena: Austrian Institute of Technology GmbH, 2013. v.1. p.1578-1586.
- Morel, R.; Morel, C. Um estudo sobre a produção científica brasileira, segundo os dados do Institute for Scientific Information (ISI). *Ciência da Informação*, v.6, n.2, p.99-109, 1977.
- Mugnaini, R. 40 anos de bibliometria no Brasil: da bibliografia estatística à avaliação da produção científica nacional. In: Hayashi, M.C.P.I.; Leta, J. (Ed.). *Bibliometria e cientometria: reflexões teóricas e interfaces*. São Carlos: Pedro e João Editores, 2013. p.37-58.
- Mugnaini, R.; Sales, D.P. Mapeamento do uso de índices de citação e indicadores bibliométricos na avaliação da produção científica brasileira. In: Encontro Nacional de Pesquisa em Ciência da Informação, 12., 2011, Brasília. *Anais...* Brasília: Thesaurus, 2011. v.12. p.2361-2372.
- Mugnaini, R. *et al.* Comunicação científica no Brasil (1998-2012): infraestrutura nacional e internacionalização. In: Congreso Internacional de Información INFO'2014, 13., 2014, Havana. *Anais...* Havana: Consultoría BioMundi/IDICT, 2014. 1 CD-Rom.
- Packer, A.L. Os periódicos brasileiros e a comunicação da pesquisa nacional. *Revista USP*, n.89, p.26-61, 2011.
- Packer, A.L. SciELO citation index no Web of Science. *SciELO em Perspectiva*. São Paulo, 27 fev. 2014. Disponível em: <<http://blog.scielo.org/blog/2014/02/28/scielo-citation-index-no-web-of-science/>>. Acesso em: 8 ago. 2014.
- Packer, A.L.; Meneghini, R. Learning to communicate science in developing countries. *Interciencia*, v.32, n.9, p.643-647, 2007.
- Pikic, A.; Mayer, M.; Macan, B. (Inter) national orientation of croatian Social Sciences and Arts and Humanities Journals Indexed in the Web of Science Database. *Društvena Istraživanja-Casopis za Opca Društvena Pitanja*, v.21, n.2, p.505-521, 2012.
- Purnell, P.J.; Quevedo-Blasco, R. Benefits to the Spanish research community of regional content expansion in Web of Science. *International Journal of Clinical and Health Psychology*, v.13, n.2, p.147-154, 2013.
- Quevedo-Blasco, R. Revistas iberoamericanas de psicología indexadas en el Journal Citation Reports de 2011. *Revista Mexicana de Psicología*, v.30, p.1-10, 2013.
- Rojas-Sola, J.I.; de San-Antonio-Gomez, C. Análisis bibliométrico de las publicaciones científicas colombianas en la categoría engineering, multidisciplinaria de la base de datos Web of Science (1997-2009). *Dyna*, v.77, n.164, p.9-17, 2010.
- Souza, E.P.; Paula, M.C.S. Qualis: a base de qualificação dos periódicos científicos utilizada na avaliação Capes. *Infocapes*, v.10, n.2, p.7-25, 2002.
- Testa, J. The globalization of Web of Science. *Thomson Reuters Expert Essays*. 2011. Available from: <http://wokinfo.com/products_tools/multidisciplinary/webofscience/contentexpansionessay/>. Cited: Jan. 12, 2014.
- Thomson Reuters. *Science Watch*. 2013. Available from: <<http://archive.sciencewatch.com/about/met/>>. Cited: Dez. 22, 2013.
- Thomson Reuters. *Web of Science*. 2014. Available from: <http://images.webofknowledge.com/WOKRS513R8.1/help/pt_BR/WOK/hs_research_tools.html>. Cited: Aug. 8, 2014.
- Wolfram, D. An analysis of canadian contributions to the Information Science Research Literature: 1989-2008. *Canadian Journal of Information and Library Science*, v.36, n.1, p.52-66, 2012.

Visibilidad de los estudios en análisis de redes sociales en América del Sur: su evolución y métricas de 1990-2013¹

Visibility of studies in social network analysis in South America: Its evolution and metrics from 1990 to 2013

Adilson Luiz PINTO²

Audilio GONZALES-AGUILAR³

Resumen

El estudio visa analizar la productividad científica de los países de América del Sur sobre el tema de los Análisis de Redes Sociales, según la representación de los países, instituciones, investigadores y referencias teóricas (las citas utilizadas en estas producciones). Estudio temporal de 1990 hasta 2013 (31/5/2013); descriptivo bibliográfico por averiguar el estado de las Análisis de Redes Sociales en contenidos de la literatura blanca; cuantitativo por aplicar volúmenes de datos de frecuencia de publicación y de citas utilizando técnicas de la Bibliometría y Cientometría; la coleta de datos fue total en la base de datos *Web of Science*, refinando los datos para los países seleccionados Argentina n=166, Brasil n=775, Bolivia n=28, Chile n=179, Colombia n=78, Ecuador n=19, Paraguay n=8, Perú n=33, Uruguay n=21, Venezuela n=48. Los países Guyana Francesa, Islas Malvinas, Suriname no publicaran ningún artículo recuperando n=922 artículos científicos; el tratamiento de los datos fue en *Statistical Package for the Social Sciences*, *CiteSpace* y *DetDraw*. Como datos relevantes, las principales relaciones por países fueron en especial con los Estados Unidos de América, cooperados con Brasil, Argentina, Chile, Venezuela y Perú. Entre los países de América del Sur la centralidad de la red se hace por Brasil, Argentina y Chile. Entre las universidades, las con más destacados son: *Universidade de São Paulo*, *Universidade Federal de Minas Gerais*, Universidad de Buenos Aires, Pontificia Universidad Católica de Chile y *Fundação Oswaldo Cruz*, donde todas estas instituciones concentran un universo de colaboraciones y hacen con que los estudios de análisis de redes sociales se desarrollen, teniendo como principales representantes Herrmann HJ *Eidgenoessische Technische Hochschule* - Switzerland, Carrasco JA *Universidad de Concepción* - Chile, Bastos FI *Fundação Oswaldo Cruz* - Brasil, Costa LD *Universidade de São Paulo* - Brasil, Moll J *Labs D'or Hospital Network* - Brasil, Benevenuto F *Universidade Federal de Ouro Preto* - Brasil, Gonzalez MC *Universidade Federal de Sao Paulo* - Brasil, Miller EJ *University of Toronto* - Canadá, Zanette DH *Centre Atomic Bariloche* - Argentina. Para la red de citas, los únicos latinos con destaque fueron Benevenuto F, Castellano C, Lancichinetti A. Sin embargo, las grandes referencias del área fueron Freeman LC, Granovetter M, Wasserman S, Scott J, Borgatti SP y Burt RS. En estudios de análisis de red social aplicado a bibliometría y cientometría los autores que publican no son identificados como los más citados y con constancia en la red. El Análisis de Redes Sociales en América de Sur trabaja directamente con el soporte de red social, el capital social, la epidemiología, la salud mental, la depresión, la violencia. Los estudios están vinculados a la medicina y enfermería, psicología, sociología, negocios y economía, física y ciencia de la información.

Palabras clave: Análisis de redes sociales. América del Sur. Bibliometría. Cientometría.

¹ Trabajo presentado en el VII Seminario Internacional sobre Estudios Cuantitativos y Cualitativos de la Ciencia y la Tecnología "Prof. Gilberto Sotolongo Aguilar" en XIII Congreso Internacional de Información - INFO'2014. Habana, Cuba.

² Universidade Federal de Santa Catarina, Departamento de Ciência da Informação. *Campus* Universitário Reitor João David Ferreira Lima, Trindade, 88040-900, Florianópolis, SC, Brasil. Correspondencia a nombre de/Correspondence to: A.L. PINTO. E-mail: <adilson@cin.ufsc.br>.

³ Université Paul Valéry Montpellier III, Département de Documentation. Montpellier, France.

Recibido el día 13/5/2014 y aceptado para su publicación el 9/9/2014.

Abstract

The aim of the study is to analyze the scientific productivity in South American countries related to Social Network Analysis in accordance with the representation of the countries, institutions, researchers and their theoretical references (citations used in these productions). The present study spans from 1990 to 2013 (31/05/2013). Bibliographical description was used to discover the status of Social Network Analysis in the scientific literature and the quantitative method to analyze publication and citation frequency data using Bibliometrics and Scientometrics. The data was collected from the Web of Science database, refining the data for the countries selected: Argentina n=166; Brazil n=775; Bolivia n=28; Chile n=179; n=78; Colombia n=78; Ecuador n=19; Paraguay n=8; Peru n=33; Uruguay n=21; Venezuela n=48 no articles were published from Guyana, Falkland Islands, Suriname, retrieving a total of 922 scientific articles. The Statistical Package for the Social Sciences, DetDraw, and CiteSpace were used for the treatment of data. Data showed that Brazil, Argentina, Chile, Venezuela and Peru maintained an especially close relationship with the United States of America. Among the South American countries, the centrality of the network consisted of Brazil, Argentina and Chile. The most outstanding universities were: Universidade de São Paulo, Universidade Federal de Minas Gerais, Universidad de Buenos Aires, Universidad Católica de Chile and Fundação Oswaldo Cruz, of which all focus on cooperation and enable studies of social network analysis to develop, with the main representatives being: Herrmann, HJ Eidgenössische Technische Hochschule - Switzerland; Carrasco, JA Universidad de Concepción - Chile; Bastos, FI Fundação Oswaldo Cruz - Brazil; Costa, LD Universidade de São Paulo - Brazil; Moll, J Labs D'or Hospital Network - Brazil; Benevenuto, F Universidade Federal de Ouro Preto - Brazil; Gonzalez, MC Universidade Federal de São Paulo - Brazil; Miller, EJ University of Toronto - Canada; Zanette DH Centre Atômico Bariloche - Argentina. As for the citation network, the outstanding Latino scholars were Benevenuto, F, Castilian, C, Lancichinetti, A. However, the field counts on the great contributions of Freeman, LC, Granovetter M, Wasserman, S, Scott, J, Borgatti, SP and Burt, RS. In studies of social network analysis applied to bibliometrics and scientometrics authors who have published are not identified as the most constantly cited ones in the network. Social Network Analysis in South America directly deals with social network support, social capital, epidemiology, mental health, depression, violence. The studies are related to medicine and nursing, psychology, sociology, business and economics, physics and information science.

Keywords: Social network analysis. South America. Bibliometrics. Scientometrics.

Introducción

La Red Social es la forma de representar las relaciones y las cooperaciones afectivas o profesionales entre integrantes que se conectan horizontalmente. Este fenómeno se manifiesta en forma de malla de múltiples hilos, que puede esparcirse indefinidamente en todos los sentidos, sin que ninguno de sus nodos pueda considerarse principal o central, ni representante de los demás. No hay un "comandante", lo que hay es una voluntad colectiva de realizar y alcanzar determinados objetivos en conjunto.

Durante las últimas décadas, el concepto de red social y los análisis de estas relaciones han sido desarrollados como una de las vías más prometedoras para medir la estructura social de la cooperación científica. Su fundamentación, sin embargo, comenzó a ser utilizada por Moreno (1934) en la consolidación de la Sociometría, al introducir la teoría matemática de gráficos, visando el dominio de la Teoría del Equilibrio Estructural.

En paralelo, Heider (1946) concluyó un estudio de análisis de equilibrio en las relaciones, partiendo de una experiencia cognitiva, determinando dos tipos posibles de cooperación: por simpatía o positiva, cuando

las relaciones son obvias dentro de un universo científico; y la relación por desaprobación o negatividad, que en general es formada por causalidades, tratándose de una función que no posee una centralidad.

A partir de estos marcos iniciales, fue determinado el origen de la Teoría de Redes Sociales, desarrollada a priori por la Universidad de Manchester en la década de 1950. Al frente de este proyecto estaban Max Gluckman y John Barnes los exponentes más importantes en la Teoría de las Redes Sociales.

En su fundamentación antropológica se visualizaron estudios centrados en la parte clínica de la Medicina y su dominio (red de análisis de enfermos y del control de prontuarios), tanto enfocados hacia la normalización de las técnicas de tratamiento en un grupo de personas, o bien partiendo de presuposiciones cognitivas y dirigidas hacia el control de tratamientos en escalas de redes de pacientes (Litwin, 1997). Otra vertiente habitual es la estructuración y la aplicación de las redes en la actividad empresarial, principalmente para el control de grupos y de producción, en procedimientos y *clusters* empresariales (Granovetter, 1985). Estas visiones son derivadas de los estudios iniciales sobre los colegios invisibles y sobre el capital social, que siempre actuaron

bajo el prisma de relación entre las personas o instituciones.

Ante este postulado, es esencial observar que las Redes Sociales no son solamente evolutivas. Se trata de una disciplina que genera nuevos conocimientos con la perspectiva de resolver problemas recientes en el área de las Ciencias Sociales, trabajando con acciones estructuradas y de pares micro y macro, explorando fundamentalmente la generación "cuantitativa" de los datos, proceso de la fundamentación de la Sociometría y la teoría de gráficos.

La utilización de generar una Red Social enfocada a los estudios de carácter métrico, utilizando la estructuración de las coautorías para formar las relaciones de cooperación científica a nivel individual, grupal y en red de autores es una constante en la actualidad. Resaltando este aspecto cuantitativo de las Redes, es esencial que los Estudios Métricos sean adoptados para poder fundamentar otros aspectos dentro de las relaciones, como las densidades, las intermediaciones y las cercanías, que son cálculos que determinan el grado de la centralidad existente en las relaciones. Esto fortalece la visión de que las Redes Sociales se fundamentan también en la Estadística, puesto que son una suma de cooperación entre dos o más puntos, o simplemente porque constituyen una representación de frecuencia científica diseñada en un mapa por inferencias similares (de autores, temáticas, revistas, documentos o instituciones científicas).

Las Redes Sociales determinan diversos comportamientos (en lo que concierne a las facetas de la antropología) y cuentan con diversos movimientos que basaron su modelo actual. Dentro de la teoría, que constituyeron las ideas de las Redes Sociales, se puede destacar dos tipos de aplicaciones.

La primera es el capital social, que comenzó a ser usado para describir una aplicación para la administración de negocios, con sus principios antropológicos hacia el relacionamiento interno de las empresas. Sin embargo eso puede ser utilizado para saber la subjetividad de la economía y política, también para vislumbrar la simbología del capital económico, su cultura y su acción social como aporte a las oportunidades laborales, al desarrollo de las profesiones y a la gestión de contactos (Bourdieu, 1977); su relación

directa entre el capital económico (infraestructura, financiación), el capital humano (educación) y el capital social (relaciones de confianza), donde esa acción hace que la unión entre el capital económico y el capital humano se potencie, en la medida en que aumenten las relaciones de confianza y la cooperación en la comunidad, y; finalmente, el capital social muestra características de organización social tales como la confianza, las normas y las redes que pueden mejorar la eficiencia social al facilitar acciones coordinadas, muy aplicadas cuando quieren saber comportamientos de costumbres.

Por otro lado, existen estudios de Análisis de Redes Sociales (ARS) basado en los colegios invisibles enfocados en el agrupamiento los actores, teniendo un aporte interesante para el medio científico en el intercambio de informaciones y experimentos. La característica de esos colegios es la interacción y la comunicación dentro de una red que funciona como un fórum educativo. En este sentido, beneficia tanto a los científicos expertos, como a los nuevos científicos que comienzan a trabajar en un determinado tema y necesitan informaciones aplicables, de algunas técnicas o de fundamentación teórica. Estas vertientes han hecho con que el medio científico comenzase a estudiar el comportamiento de estas sociedades, basando el aporte para vislumbrar el análisis de las relaciones oscuras o poco visibles.

Actualmente, para analizar los colegios se emplea la representación gráfica basada en las Matemáticas de características cartesianas, que se conoce como Sociometría, cuya vía para la ciencia se fundamenta en el momento cuando se evalúa el todo del conjunto por sus partes. Se trata, por lo tanto, de una ciencia social que se ocupa del individuo como integrante de un determinado grupo, y de las relaciones con sus pares. Esta fundamentación estructural puede ser representada mejor por sus modelos: como el de *Estrella* que está compuesta de diversas líneas apuntando hacia un punto céntrico, que representa un individuo envuelto en la relación con cada uno de los demás participantes, y donde las flechas representan relaciones tales como reciprocidad, ignorancia o rechazo; el de *Corriente* que se caracteriza por una serie de relaciones en fila, con un intermediario que hace la función de enlace de los pares;

el de *Pares*, compuesto por una relación recíproca entre dos individuos; el de *Poder*, representado como una atracción entre grupos distinguidos que se funden por iniciativa de uno de los individuos, y; el de *Aislamiento*, determinado por un asunto no escogido por demasiados individuos, cuya interacción social no está consolidada.

El modelo de la Sociometría de Jacob Moreno puede ser aplicable en cualquier estudio relacionado a las Redes Sociales, mientras que sus representaciones pueden ser mejor identificadas si se les combina con el análisis de autores o instituciones, para una mejor visibilidad gráfica.

Como aporte de estudio, indagamos cómo se comporta el análisis de redes social en América del Sur según su productividad científica de impacto; visando evaluar y analizar la productividad científica de estos países sobre el tema de las ARS, fundamentado en (i) evaluar el grado de centralidad de los países en el tema; (ii) analizar el comprometimiento de las instituciones; (iii) identificar los investigadores más representativos en producción, y; (iv) vislumbrar el comportamiento de los referenciales teóricos en el tema.

Materiales y métodos basados en los estudios métricos de la información

A principios del siglo XX se plantearon los primeros estudios bibliométricos. Evaluando las frecuencias, distribuciones y probabilidades de la información científica y tecnológica, pasando a nuevas métricas, en especial por la dimensión de los estudios, como el ejemplo de la Cientometría.

Algunos estudios se propusieron atender a la capacidad estadística de la producción científica y tecnológica. En uno de estos estudios se establecieron los fundamentos de la ley del cuadrado inverso (Lotka, 1926), afirmando que el número de autores que hacen "n" contribuciones en un determinado campo científico es aproximadamente $1/n^2$, mientras los que hacen una sola contribución es de más o menos el 60%, que se ha utilizado posteriormente para medir el grado de importancia de los autores en sus disciplinas y sus correlaciones con otros autores, en grado de proximidad.

Por la misma época, otro estudio se preocupó en identificar los métodos para localizar las revistas-núcleo, conociendo de esta forma los principales países investigadores y el desarrollo de estudios a través de citas, a la vez que realizaba un análisis de revistas claves y daba origen a la medida de la ciencia (Gross, P.L.K. & Gross, E.M. 1927).

Posteriormente, Bradford (1934) propuso el método para conocer los autores más citados y en qué revistas eran citados, para lo que utilizó un estudio muy semejante, en el que trató de representar su teoría sobre el núcleo de revistas, dividiendo la suma total de las revistas por tres partes, teniendo así sus núcleos de característica 1, 2 y 3, siendo de alta intensidad la 1, de media intensidad la 2 y de baja intensidad la 3. Ejemplo: Un total de citas recibidas por una revista es de 1500, el primer núcleo contiene las 500 primeras, el segundo las otras 500 y el tercero las 500 finales. Sin embargo, las 500 primeras estarán representadas por 55 autores, principalmente porque son los que más producen, y por este motivo quedan determinados como primer núcleo. El segundo núcleo está representado por 220 autores, con una representación en torno de dos trabajos por autor; y el último núcleo es representado por 500 autores, todos con una citación recibida.

Algunos años después, exactamente en 1949, se publicó uno de los estudios más complejos de todos los tiempos por el científico Zipf (1949), quien estableció una comparación de las palabras clave con las citas, donde observó que el uso de las palabras en cualquier lengua está claramente definido por valores constantes. Así, definió que el número Y de veces que aparece una palabra es inversamente proporcional a su rango X, es decir, $Y=a/X$. Para un estudio de análisis de citaciones, se representaría mediante una frecuencia elevada de palabras claves, donde éstas estarían entre las más citadas. Esta técnica incluso puede ser utilizada para textos aleatorios, tal como lo aplican los motores de búsqueda de Internet.

Las ideas fundadoras de disciplina bibliométrica remontan a la década de 1950, con Vannevar Bush, explorando el poder de las citas estadísticas; también con el origen del Instituto de Ciencias de la Información, donde Garfield (1955) aprovechó todas las "teorías preliminares" para desarrollar un recurso de normalización del contenido bibliográfico, aplicando una base de datos

en la que identificó cuales eran las revistas "más importantes" y cuáles eran los países con mayor representación internacional, y posteriormente implantó un análisis sobre los autores más citados. La Science Citation Index, la Social Science Citation Index, la Arts & Humanities Citation Index y el Journal Citation Report representan los productos más conocidos de su compañía.

Las iniciativas de Garfield no fueron inicialmente una acción en favor de la ciencia, puesto que tenía por detrás intereses financieros, pero sirvió de inquietud del potencial del recurso de las bases de datos, pudiendo también mirar el análisis en las citas, teniendo como comparar las citaciones con el desarrollo de los equipos de investigación y el crecimiento exponencial de la ciencia (Price, 1964), determinando la aplicación de la ciencia por la ciencia.

De las iniciativas en bases de datos, Kucera y Francis (1967) desarrollaron un análisis computacional para eliminar las palabras vacías, a partir de una lista de palabras vacías generada mediante un diccionario de frecuencias. El análisis demostró la naturaleza de las diferencias sintácticas entre los distintos géneros de la escritura representados en la base de datos, especialmente entre la prosa informativa y la prosa imaginativa. Esta técnica pasó a incorporarse a los estudios métricos, como consecuencia de haber fundamentado algunos programas de control sistémico de las palabras, introduciendo los estudios que se basan en los términos más utilizados.

En la década de 1970 los estudios métricos sufrieron un cambio en la forma de su aplicación, con mayor interés a los métodos basados en la co-ocurrencia de las palabras, utilizando las bases de datos más representativas en la época, como el Social Science Citation Index y Science Citation Index. Siendo posteriormente visto modelos de análisis a los mapas de co-citaciones, teniendo como base los programas Pajek <<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>> y Ucinet <<https://sites.google.com/site/ucinetsoftware/home>> para representar los análisis de redes.

La gran preocupación existente por realizar estudios métricos para cuantificar la progresión en la

ciencia y en la tecnología ha hecho que muchas aplicaciones prácticas de las teorías sobre medición se puedan utilizar, pero todo basado en tres aspectos: la actividad, la productividad y el progreso.

Ante estos tres aspectos de la ciencia, la Organización para Cooperación Económica y Desarrollo (Organization for Economic Cooperation and Development, 1975) formuló un manual práctico para el desarrollo de la medición de las actividades científicas, convirtiéndose en una referencia para las encuestas de Innovación y Desarrollo de sus países miembros. A partir de esta iniciativa, la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) propuso la norma para las encuestas de Innovación y Desarrollo en todos los países del mundo, estableciendo las formas de investigación: básica, aplicada y de desarrollo experimental.

Frame *et al.* (1977) hicieron un trabajo en el que analizaron la literatura de economía con cobertura internacional, preocupándose mucho por la visibilidad del Este europeo pero también, como casi todos los análisis métricos en bases de datos, utilizaron ISI y Medline⁴.

La medición de la ciencia y la tecnología pasó por una época de preocupación creciente respecto a la política aplicada a sus desarrollos, como demuestra que los principales países empezasen a confeccionar propuestas para una autosuficiencia en nuevas tecnologías, visando los recursos de informática, como la actualidad. Pese a que los estudios de producción científica y colaboración entre autores fueron basados en este principio, para poder ganar espacio en el medio científico, sin embargo con limitaciones, puesto que no contaban con recursos apropiados para continuar con el desarrollo que la ciencia y la tecnología entonces requerían.

En el inicio de la década de 1980 se empezó a discutir el progreso tecnológico desde su derivación de las aplicaciones científicas del mundo académico, resaltando los factores culturales, sociales y económicos determinados por el progreso de la ciencia, y viéndose cómo el progreso tecnológico antecede y estimula a la actividad científica.

⁴ Servicio de Información para la Salud de la *National Library of Medicine* y de la *National Institutes of Health*.

En paralelo los gobiernos comenzaron a invertir en los indicadores de input y output, principalmente en las instituciones públicas de enseñanza e investigación, desarrollando para cada país un modelo de análisis de los datos, denominados de informes de ciencia y tecnología. Sin embargo, esta actitud obtuvo muchísimos errores de datos, principalmente porque algunos países tendrían la necesidad de justificar sus gastos, y algunos tuvieron que maquillar sus datos.

Una de las soluciones para poder identificar realmente los gastos de los gobiernos fue la distribución de publicaciones por área de la ciencia, a través del principio de autoorganización y agrupación de vectores n-dimensionales en un espacio bidimensional. Este estudio consistía en demostrar que una información de entrada por sí sola a través de una estructura propia y una descripción funcional del comportamiento de la red puede formar mapas de características de citas, de topologías, o de autoridad. Estos mapas tienen como característica organizar la información de entrada clasificándola automáticamente, y el sistema de autoorganización básico tiene la función de agrupar la frecuencia conforme sus características y relacionarla con otras temáticas por proximidad (Kohonen, 1982).

Otra iniciativa, utilizando la autoorganización de los mapas, fue de Honkela *et al.* (1996), donde aplicaron un nuevo método, con una función en la web de los mapas de autoorganización, representando documentos por sus semejanzas y sus características comunes de campos, principalmente cuando existe un standard en los puntos que distinguen las temáticas del texto. El método completo implica una arquitectura de dos niveles: por un lado están los mapas de autoorganización y por otro lado están los mapas de categorías de las palabras y los mapas de documento, así como los medios para la exploración interactiva de la base de datos. Volviendo la versión web mucho más dinámica y completa.

El método de mapas de autoorganización es ideal para analizar materias dentro de un universo de palabras, diferenciadas por nombres y colores (por defecto). Puede ser utilizado también para identificar grupo de autores que trabajan juntos o que lo hacen dentro de una materia, sirviendo incluso para evaluar una posible red

en tres dimensiones, como algunos softwares ya lo hacen, por ejemplo el Gephi <<https://gephi.github.io/>>.

Otro trabajo clásico, pero con un carácter más enfocado hacia la recuperación de la información, fue el de Salton *et al.* (1982), por el que determinaron el principio del Modelo del Espacio Vectorial (MEV) que posteriormente fue descrito como aportación fundamental para el desarrollo de la Recuperación de Información, junto a soportes de tareas como la desambiguación del sentido de las palabras y la clasificación de los textos. Partiendo de su aplicación conceptual a la recuperación de la información, es posible conceder el valor más alto a las palabras que causan la mayor diferenciación entre los documentos de una colección que se pretende indizar, relación luego muy utilizada en los análisis bibliométricos, principalmente porque buena parte de estos trabajos son extraídos de bases de datos y pasan por un proceso de recuperación de información y después por los de tratamiento y normalización.

Casi al mismo tiempo Cronin (1984) discutía las causas por las que se citaba a un autor y no a otro, como una forma de "ideología científica" que las revistas solían hacer cuando se pretende publicar un determinado trabajo y no se refería a los autores "clásicos". Sin embargo, las citas siempre son una influencia del área de conocimiento, en especial para las áreas de gran impacto científico; y dentro de ese impacto es posible determinar los estudios más relevantes, como siendo teóricos, independiente si la estructura corresponde a métodos o aplicaciones.

Para este tipo de metodología es posible producir normalización basada en la función de tipificación utilizando el factor de impacto, trabajando con variaciones al tiempo que sus escalas de diferentes categorías son homogéneas. Este tipo de aplicación puede servir como referencia a la hora de situar la posición del dominio de áreas, a diferencia de otros cálculos en los que el valor resultante se sitúa dentro de un rango.

White y McCain (1989) propusieron un análisis donde las métricas de las cocitaciones son un indicador muy relevante para las relaciones de los autores. Consiguientemente observaron que cuanto más alta era la cocitación entre los pares de los autores, más fuerte

eran sus relaciones según lo percibido por citers, y más fácilmente interpretables los mapas basados en ellos.

Schubert *et al.* (1989) presentaron una posible manera de caracterizar la importancia relativa de los enlaces de una institución con respecto a otra. Se trata de un índice de coautoría asimétrica, utilizada posteriormente en los estudios de Redes Sociales, donde ellos aplicaron como un comparativo del *Science Citation Index* con el Ulrich's International Periodicals Directory.

A principios de los años 1990, los documentalistas observaron que los autores más representativos en la temática empezaban a aplicar una forma de consumo de información en los centros de documentación y bibliotecas para control interno de préstamos y consultas. Pero aplicaban tan solo una décima parte de las técnicas métricas, y en muchos casos de forma equivocada de manera que el análisis científico no era muy bien empleado por los profesionales de la información, mientras que los grandes trabajos de esta década aplicados a la Biblioteconomía y Documentación seguían en manos de químicos, físicos, ingenieros, científicos de la salud y biólogos.

Dentro del universo académico el consumo de información es denominado como análisis de citas, pudiendo ser clusters de citas, visando identificar cuáles son las materias más estudiadas, recuperando por un lado parte de la bibliografía internacional en bases de datos y por otro lado identificar la frecuencia científica de algunos autores en las memorias científicas.

Egghe y Rousseau (1991), presentaron un artículo sobre la noción de concentración que puede estudiarse aplicando un número de principio de transferencia. Los principios de transferencia implican otros requisitos naturales de concentración, entre los cuales se sitúa el principio del aumento nominal. Como parte práctica, utilizaron una comparación entre el principio de transferencia de Dalton (el principio de suma de cantidades fijas) y la orden de dominación de Lorenz (la proporción del área entre la curva), distinguidos como principios simples y fuertes en la transferencia, conduciendo a una clasificación de las medidas de la concentración, por la que se identificó que el coeficiente de variación y otras medidas derivadas son las más adecuadas para este tipo de evaluación.

En la misma época Frakes (1992), se introdujo en un estudio desde los algoritmos en el que empleó muchos sistemas propios del análisis de proceso y de la lingüística del texto para estudiar las búsquedas y la recuperación en bases de datos. El contexto científico se situó en el efecto de dinamizar el funcionamiento de la recuperación desde experiencias previas y con precisión, siendo posteriormente utilizado por otros autores en la comprensión de los índices.

Con eso, fue posible destacar la relación existente entre las dimensiones de la ciencia y de la tecnología, donde situaron una de las características principales de los sistemas de innovación; pero la ciencia es el hilo para el progreso tecnológico.

Siguiendo esta teoría, tenemos como base las búsquedas de los esfuerzos en Ciencia y Tecnología (CyT) por medio de indicadores de input como volúmenes de inversión en investigación científica y tecnológica. Y que tal configuración fue el resultado de la evolución respecto a la manera de entender la cadena de producción del conocimiento científico y tecnológico, generando los outputs necesarios al desarrollo en CyT.

Como complemento, tenemos el factor de impacto de las revistas de los paquetes del Institute for Scientific Information, basado en evaluar la cantidad de citas en dos años, dividida por el número de artículos publicados en el mismo período en una determinada revista, como una fórmula aritmética.

Gibbs (1995) realizó un informe sobre la salida masiva de científicos de sus países de origen, generalmente en desarrollo, para buscar oportunidades en países ya desarrollados, principalmente en Estados Unidos (EEUU) lo que supuso aún más subdesarrollo para los países emergentes. Esto provocó que los bajos índices productivos se hiciesen casi constantes.

Posteriormente, Narin *et al.* (1997) buscaron las evidencias empíricas en el vínculo creciente entre la ciencia - fundamentalmente financiada por el sector público - y la capacidad innovadora del sector industrial, pero aplicada a países desarrollados. Sin embargo, la inquietud de la comunidad científica mundial se figuró en las funciones de la publicación científica, en especial para el progreso de la ciencia como un todo.

A partir del año 2000 los estudios enfocados hacia la ciencia iban logrando mayor impacto en la sociedad, por lo que muchos países comenzaron a destacarse a nivel mundial, principalmente los llamados emergentes. Sin embargo, buena parte de esta representación se debió a la internacionalización de las bases de conocimiento (ISI, Pascal, Pro-Quest entre otras) y de desarrollo locales, que buscaban atender a esta necesidad.

En ese mismo año fue muy relevante la aplicación del factor de impacto en bases de conocimiento, en especial por el estudio de Amin y Mabe (2000) que hicieron un repaso de esta técnica de medición de las revistas, pero visando las limitaciones de este tipo de indicador, pues todos los recursos son determinados por sus puntos fuertes y sus debilidades.

Un año después, White (2001) fundamentado en las citas, cocitas y coautorías, hizo resurgir las técnicas de la visualización de la información de manera automática, generando perfiles personales de autoría a partir de los datos contenidos en las bibliografías (los conocidos Cameo's). Este recurso es primordial para la recuperación de los documentos originales. Hay que resaltar que una preocupación constante de White consistió en observar la teoría de la autocitación, recurso indebidamente utilizado por algunas autoridades en las bases de datos internacionales.

La evolución científica sugiere la necesidad de fomentar esfuerzos para encontrar y desarrollar nuevas formas de publicación de las monografías. Thompson Reuters (2002) hizo un examen de 6.708 citas de monografías y de artículos de revista sobre literatura británica y americana del siglo XIX, demostrando que los eruditos de la humanidad recuperan una gran cantidad de fuentes primarias y secundarias con una edad de vigencia de amplio espectro. Las citas revelaron una gran confianza en las citas de las monografías, y también se identificaron una gran cantidad de revistas clave.

En 2003, un trabajo de Chaomei Chen fijó convincentemente las teorías generales referentes a la cartografía y a la comunicación visual de la ciencia, teniendo como fundamento el marco de *Generalized Similarity Analysis* (GSA) (modelo para extracción de relaciones semánticas entre documentos y enlaces hipertextuales). El objetivo era desarrollar técnicas de

visualización de la ciencia, en forma de mapas de conocimiento. Posteriormente a esa publicación Chen desarrolló un programa denominado *CiteSpace* para establecer las relaciones científicas que se daban entre autoridades.

Todos los estudios anteriormente citados, ya clásicos y consagrados, sirvieron de soporte para que Freeman (2004) pudiera abordar una temática que ya empleara en la década de 1930 Jacobo Moreno, la de las redes sociales. La diferencia entre el estudio de Freeman y el de Moreno fue que la representación gráfica no era el principal objetivo, sino los datos cuantitativos de las cooperaciones científicas en forma asimétrica. Del estudio se derivó de inmediato un índice social de la citación de la ciencia. Freeman solamente hizo públicas las iniciativas que estaban realizando muchos investigadores en todo el mundo respecto a las redes sociales, pero que llevaban a explicaciones distintas, incluso desde la denominación, pues se apreciaron propuestas entorno a: "relaciones científicas", "Redes Sociales" y "coautorías".

Es probable que la progresión y difusión de la bibliometría sea la evolución de la preocupación de saber lo que se investiga, los representantes de estas publicaciones, las interacciones existentes en las investigaciones, y saber de hecho en creciente número de investigadores que tenemos actualmente, en especial por los incrementos de publicaciones en las bases de datos y del incremento de investigadores en el mercado científico.

El reflejo de todo eso es la orientación de carácter internacional en el mundo científico, que se basa en bases de datos de gran amplitud, como nuestro estudio, que trabaja directamente con una base de datos de ámbito internacional, la *Web of Science*.

La Web of Science como recurso informacional para análisis de redes sociales

La *Web of Science* es una base de datos del *Institute for Scientific Information*, iniciada por un proyecto de Eugene Garfield en 1960 con la visión de indizar información científica de las revistas americanas de la

ciencia, especial las ciencias puras, reportada en la base *Science Citation Index Expanded*.

En la década de 1980, con la comercialización de esta base de datos, comienza a ser recuperada también información en las áreas de Ciencias Sociales, Humanas y Artes, a partir de las bases *Social Science Citation index* y *Arts & Humanities Expanded*.

Todo el contenido de estas tres bases de datos están desde 1945, con un plus de información, dando toda la información registrada en la base para buscar servicios, los índices de productividad, índice-h y el índice de citas de autores y de revistas, con la base *Journal Citation Report*.

Esta iniciativa fue la primera del mundo de agregar las citas como un punto de partida para verdadera acción de exploración de las ciencias en el mundo. Otro dato importante es que fue la primera base a trabajar con índices sociométricos de la ciencia, con vista a la cientografía o teoría de gráficos a la ciencia, generando mapas automáticos de las relaciones de los autores.

Más recientemente la *Web of Science* se fundió con otros paquetes, como bases de actas científicas, incorporado en la década de 2000 con la compra de este repertorio por la *Thomson Reuters*.

La *Thomson Reuters*, ha mirado una oportunidad de controlar información científica creando, a partir de la *Web of Science*, otra plataforma denominada *Web of Knowledge*, indizando las bases: *Arts & Humanities Citation Index*; *Biological Abstracts*; *Biosis Citation Index*; *CAB Abstracts*; *Conference Proceedings Citation Index*; *Current Chemical Reactions*; *Current Contents Connect*; *Derwent Innovations Index*; *Essential Science Indicators*; *FSTA Food Science Abstracts Technology*; *Salud Global*; *Chemicus Index*; *Inspección*; *HighlyCited.com ISI*; *Journal Citation Reports*; *MedLine*; *Science Citation Index Expanded*; *Social Sciences Citation Index*, y Expediente Zoológico (*Thomson Reuters*, 2010), controlando buena parte de la información académico/científica.

Como aporte informacional para la *Web of Science* en los trámites de análisis de redes sociales fueron creados por la *Thomson* en finales de 2000 un sistema basado na teoría de grafos, con representación limitada para identificar los análisis egocéntricos de autoridades

y revistas. Sin embargo, actualmente este recurso no es más visible en el sistema.

La alternativa para eso fue la versión actualizada del *CiteSpace*, donde Chen (2003), ha agregado contenidos para tratar la información por acoplamiento bibliográfico, un recurso muy significativo para trabajar el análisis de redes sociales ARS en contenidos científicos, donde a continuación iremos representar por gráficos la producción científica de ARS en los países de América del Sur.

Recientemente, nuevos sistemas de indexación de citas se han lanzado para competir con *Web of Science (Thomson-Reuters)* por ejemplo, la *Scopus*, el sistema apoyado en Elsevier. Sin olvidar *Google Scholar* que se interesa a las redes de citas de la Web. Desde el inicio de la web hace casi dos décadas se han propuesto indicadores de impacto similares a los utilizados en la cienciometría para los sitios web institucionales y páginas web, con base en las estructuras de enlace.

Discusión

En nuestros análisis vamos inferir sobre dos puntos básicos, el primer enfocado en los análisis de redes sociales, teniendo la teoría de gráficos y la sociometría como centrales, y por otro lado los estudios métricos de la información para generar las inferencias del comportamiento de las relaciones científicas.

Como punto clave, existen las relaciones científicas por los países de América del Sur y sus colaboradores externos. De estas forma consideramos 3 países con centrales en la publicación (Brasil, Argentina y Chile), representando América del Sur y los Estados Unidos de América y España como elementos de ligación entre la Sudamérica con el mundo. Esta relación se hace en especial con Chile y Argentina. Brasil tiene toda tu rama individual (Figura 1).

La bibliometría está trabajando ahora para desarrollar modelos estadísticos para medir la evolución de la ciencia y la tecnología, y permitir la definición de los grupos temáticos que componen la realidad subyacente al avance de la ciencia en Latinoamérica.

Otro dato crucial es la inferencia de centralidad y proporción de la productividad (n=775) que representa

Brasil en el análisis, siendo seguido por los EEUU (n=209), Chile (n=179), Argentina (n=166), Colombia (n=78), España (n=71), Inglaterra (n=51) y Venezuela (n=48).

Dentro del universo brasileño, el país ha logrado relaciones aisladas con países europeos (Rusia, Italia, Austria, Suecia, Portugal, Suiza), con asiáticos (China y Japón), con América Central y Caribe (Cuba) y con África, Oriente Medio y Oceanía (África del Sur, Israel y Australia).

En la proporción científica de los países también es relevante mencionar la importancia que instituciones representan en los análisis, como el grado de productividad de la *Universidade de São Paulo*, *Universidade Federal de Minas Gerais*, Universidad de Buenos Aires y la Pontificia Universidad Católica de Chile. En una segunda escala tenemos *Universidade Federal de Santa Catarina*, *Universidade Federal de Rio Grande do Sul*, *Fundação Oswaldo Cruz*, representados en la Figura 2.

Las áreas más recurrentes de conocimiento en el ARS fueron las de Medicina (Salud Pública, el Medio Ambiente y la Salud Ocupacional con 170 estudios;

Enfermería n=51; Psicología n=38), Ciencia de la Información (n=43) y de la Computación (Sistemas de información n=30, Computación n=28), la Física (n=73) y las Ciencias Administrativas (Administración n=34; Negocios n=26; Ciencias Económicas n=33).

En representación a estas instituciones y sus áreas de investigación, algunos autores se destacan por su forma de cooperar y de producir, como fue el caso de Herrmann HJ *Eidgenoessische Technische Hochschule* - Switzerland, Carrasco JA Universidad de Concepción - Chile, Bastos FI *Fundação Oswaldo Cruz* - Brasil, Costa LD *Universidade de São Paulo* - Brasil, Lind PG *Universidade Federal do Ceará* - Brasil, Moll J *Labs D'or Hospital Network* - Brasil, Benevenuto F *Universidade Federal Ouro Preto* -

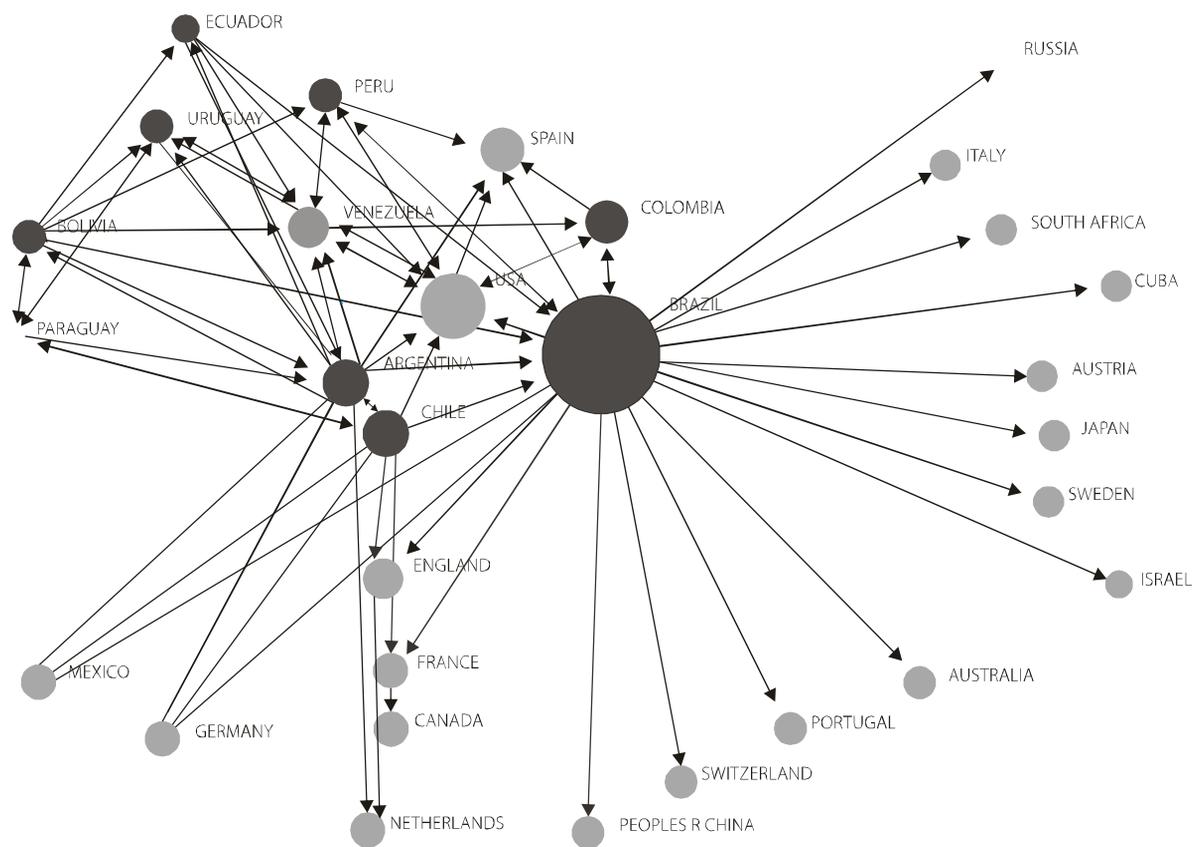


Figura 1. Relaciones de los países Sudamericanos en Análisis de Redes Sociales.

Fuente: Datos de la investigación (2014). Generado por los autores en *NetDraw*.

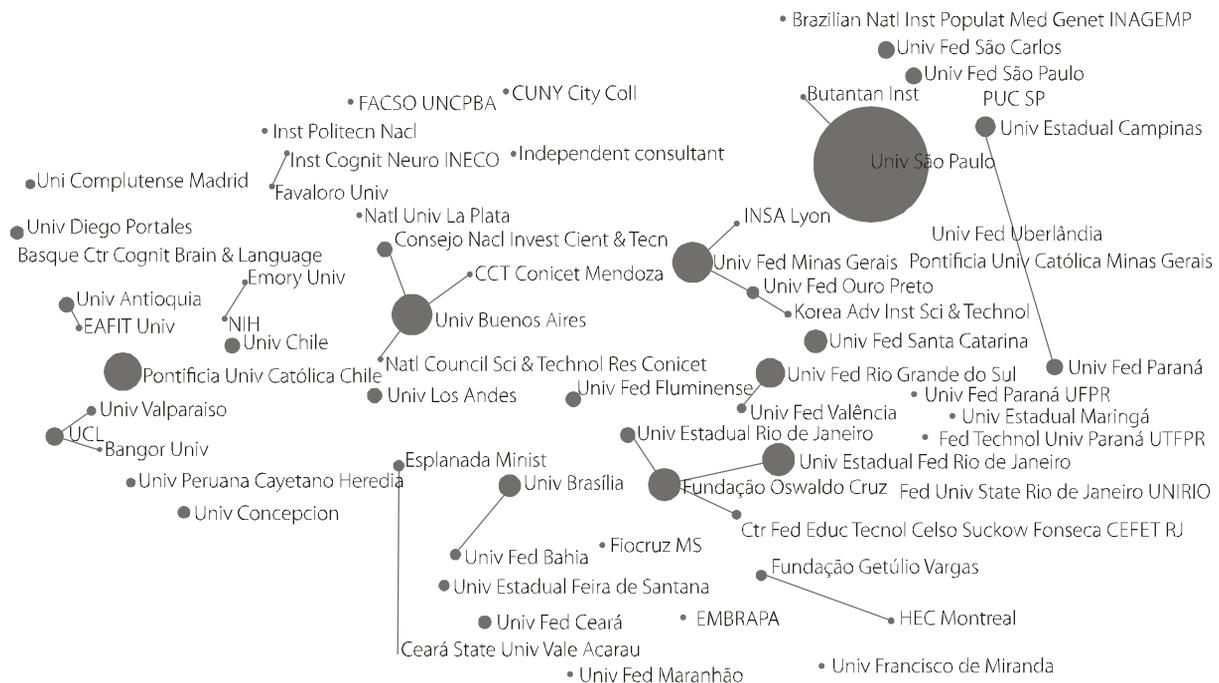


Figura 2. Comportamiento de productividad de las Instituciones en Análisis de Redes Sociales de Sudamérica.

Nota: CCT: Centro Científico Tecnológico; CUNY: The City University of New York; EMBRAPA: Empresa Brasileira de Pesquisa Agropecuária; FACSQ: Facultad de Ciencias Sociales - Universidad de Chile; INSA: Institut National des Sciences Appliquées de Lyon; UCL: University College London.

Fuente: Datos de la investigación (2014). Generado por los autores en CiteSpace.

Brasil, Gonzalez MC *Universidade Federal de São Paulo* - Brasil, Miller EJ *University of Toronto* - Canadá, Zanette DH *Centre Atom Bariloche* - Argentina, Anderson CB *University of Texas* - EEUU, Andrade JS *Universidade Federal do Ceará* - Brasil y Braunstein LA *University of Boston* - EEUU.

Sin embargo, tener producción elevada y destaque en este aspecto no influencia en la centralidad de una red, donde es más importante que estos autores sean las ligaciones entre instituciones, representantes e incluso entre países.

De esta forma, la Figura 3, tiene estas dos variables, (i) con un cuadro para la importancia de la frecuencia científica de algunos autores, como mencionado anteriormente, sus publicaciones, y (ii) un gráfico para la centralidad de influencia en la red. Algunos autores consiguen tener destaque en ambas inferencias, como es el caso de Bastos FI, Uchoa E, Firmo JOA, Herrmann HJ y Zanette DH.

Curiosamente, estas frecuencias y centralidades tienen especial destaque a partir de 2006/2007, con influencia de Almeida V, seguido de Perez AM y Ford GA.

En contrapartida de la productividad y su centralidad en las publicaciones existen las influencias científicas que estos autores tuvieron para construir el conocimiento registrado.

La forma de construcción del conocimiento registro es a partir de otros autores, en forma de citas y de acoplamiento bibliográfico, muy recurrentes en ARS aplicados a las citas de los estudios bibliométricos y cuantitativos.

Una vez más, la forma de la cuantificación de frecuencia poco influencia en la centralidad y en el acoplamiento bibliográfico, pues uno es sólo representado en cómo fueron utilizados los referenciales teóricos, mientras que el otro muestra cómo algunas citas hacen toda la diferencia en contextualizar, en especial por no aportar una connotación superficial, sino específica sobre el tema abordado. En nuestro caso, estaremos con estas dos connotaciones frente a los Análisis de Redes Sociales.

Tratando de la frecuencia de citas, tenemos: Newman MEJ (n=152) seguido por Watts DJ (n=77), Barabasi AL (n=73), Wasserman S (n=59), Albert R y Galam

S (ambos con n=57), Borgatti SP y Wellman B (ambos con n=48), Stauffer D (n=45), Granovetter M y Minayo MCS (ambos con n=44), Rozzi R (n=43), Axelrod R (n=37), Scott J (n=34), Bourdieu P, Castells M y Freeman LC (todos con n=33), Friedman SR y Gulati R (ambos con n=32), Castellano C (n=31), Latour B y Sznajd-Weron K (ambos con n=30), Pastor-Satorras R y Putnam R D (ambos con n=29), Burt RS y Gonzalez MC (ambos con n=28), Bronfenbrenner U y Moll J (ambos con n=27), Amaral Ian, Coleman J S y Klemm K (todos con n=26) y Carrasco JA (n=25).

Por otro lado, tenemos el acoplamiento bibliográfico en que son vislumbradas la primera cita en

la base de datos de los trabajos en el tema Análisis de Redes Sociales de los artículos vinculados a los autores de países de América del Sur.

En la Figura 4 las flechas relacionan las formas como estas citas son conjuntamente utilizadas por los autores, donde tenemos un inicio de las mismas por Freman LC, Granovetter M, Wasserman S y Erdos P, donde estos autores fundamentan los estudios de ARS desde 1990, según las publicaciones sudamericanas.

En una segunda esfera ya aparecen los acoplamientos de estas citas con Scott J, en especial por tratarse de fundamentos de ARS y de la relación directa

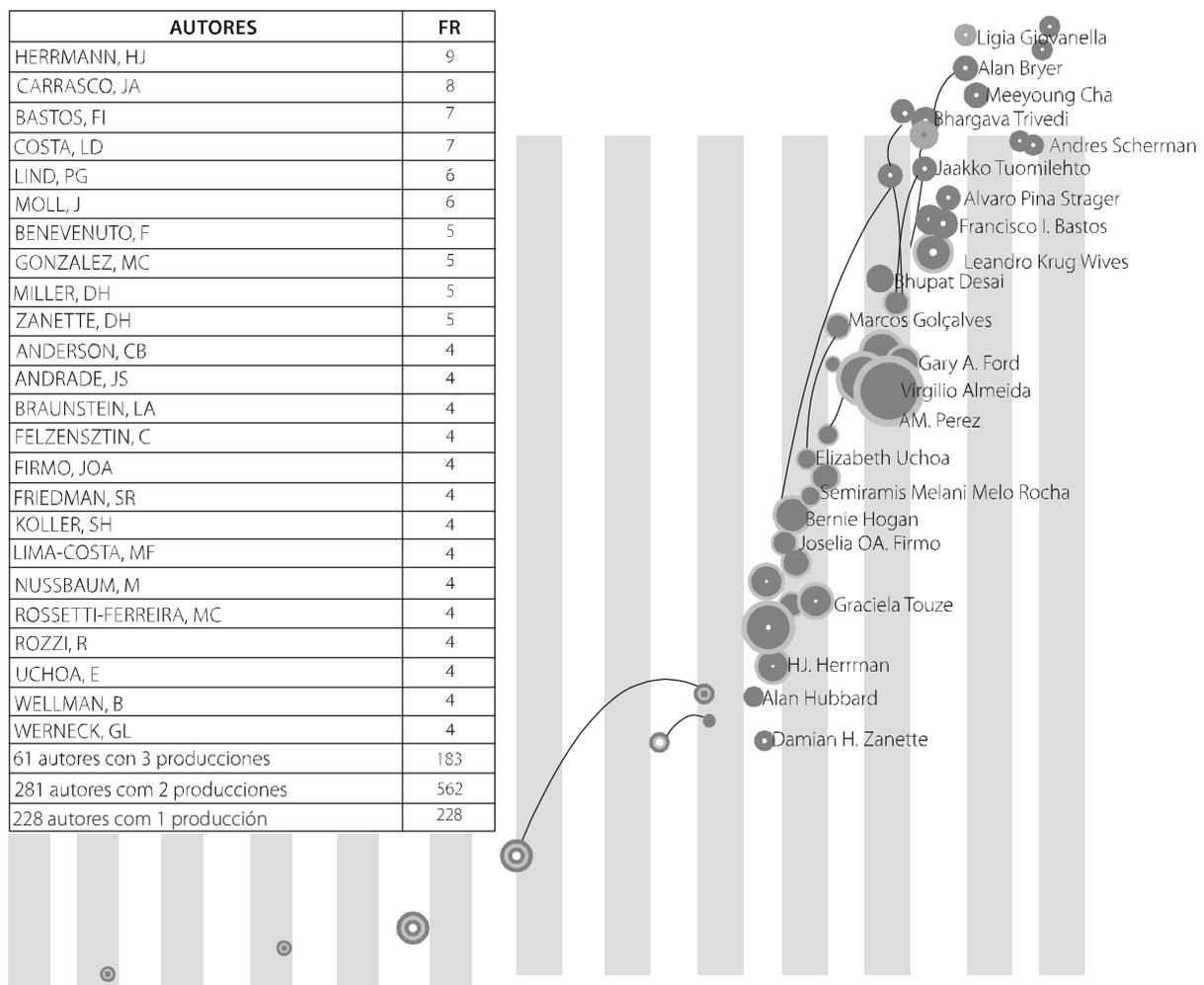


Figura 3. Frecuencia (FR) y centralidad de los autores que publican en Análisis de Redes Sociales.

Fuente: Datos de la investigación (2014). Generado por los autores en CiteSpace.

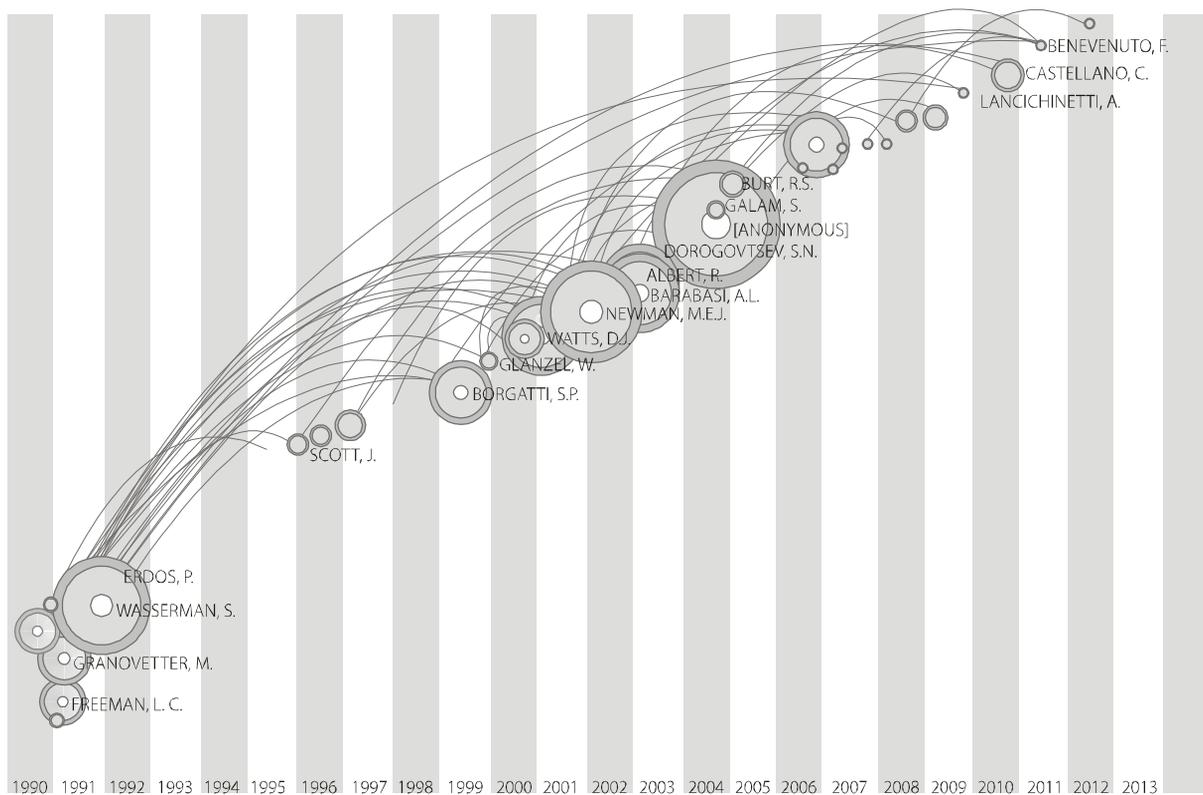


Figura 4. Citas y acoplamiento bibliográfico.

Fuente: Datos de la investigación (2014). Generado por los autores en *CiteSpace*.

que existe de las teorías de su fundamentación (Gestalt, Gráficos, Sociometría y Dinámica de los grupos).

Otros acoplamientos ocurren con más intensidad a partir de los años 2000, con la aplicación de ARS aplicados a las áreas médicas, de negocios y de la divulgación científica de los estudios métricos de la información. Esta última utiliza las ARS para justificar la inversión de recursos (*Input*) y sus colaboraciones por autoridades, instituciones y países, y los retornos de los investimentos a la investigación, en forma de artículos publicados en conjunto (*Output*). En este sentido tenemos Glanzel W y Dorogovtsev SN.

Para finalizar, la influencia latinoamericana sola, empieza a tener relevancia en las citas y en acoplamiento de las citas a partir de 2010, cuando la relación directa de Castellano C (Venezuela) y Benevenuto F (Brasil) empiezan a ganar espacio en la Ciencia de la Computación y en la Física.

Conclusión

En primer lugar, realizamos análisis de redes sociales, destinado a responder a algunas preguntas como: ¿Quiénes son los autores más importantes de la zona? ¿Existen grupos cerrados de autores? ¿Hay alguna persona en el medio de la mayoría de estos grupos? ¿Esta persona es también muy importante?

Los estudios de ARS son aplicados a diferentes necesidades y eso quedó visible en este estudio, donde tenemos las más diversificadas aplicaciones, como en Salud Pública, el Medio Ambiente y la Salud Ocupacional, Enfermería, Psicología, Ciencia de la Información y de la Computación, Física, Administración y Ciencias Económicas.

Si la bibliometría se ha impuesto en el mundo de la investigación es porque su fundamento es casi familiar en internet: la red y los análisis de redes. Más allá de los clichés a menudo reduciendo el factor de impacto y los

índices de evaluación individual, la bibliometría está trabajando ahora para desarrollar modelos estadísticos para medir la evolución de la ciencia y la tecnología.

Por otro lado, la influencia de su aplicación a los estudios métricos de la información (bibliometría y cientometría) hacen con que identifiquemos a los autores que más publican, sus instituciones de trabajo y quienes son sus referencias en las publicaciones en forma de citas y co-citaciones.

De esta forma el ARS en América de Sur trabaja directamente con el soporte de red social, el capital social, la epidemiología, la salud mental, la depresión, la violencia, basados en las principales instituciones de este continente, como *Universidade de São Paulo*, *Universidade Federal de Minas Gerais*, Universidad de Buenos Aires, Pontificia Universidad Católica de Chile, *Universidade Federal de Santa Catarina*, *Universidade Federal do Rio Grande do Sul* y *Fundação Oswaldo Cruz*, siendo cinco brasileñas, una argentina y una chilena.

Toda la publicación tiene una relación directa con sus productores personales, como las de los brasileños Bastos FI, Costa LD, Lind PG, Moll J, Benevenuto F, Gonzalez MC y Andrade JS; el argentino Zanette DH y el chileno Carrasco JA, sin embargo el análisis contó directamente con las interferencias de personajes extranjeros, como Herrmann HJ de Suiza, Miller EJ de Canadá, y de los americanos Anderson CB y Braunstein LA. En las citas la influencia fue toda extranjera hasta 2010, basado en los personajes que difundirán el ARS en el mundo y sus aplicaciones, como Freeman LC, Granovetter M, Wasserman S, Scott J, Bogatti SP, Watts DJ y Burt RS.

Los grupos estudiados nos permiten confirmar la teoría de Mullins citado por Chen, que las agrupaciones de cocitación representan de hecho la estructura intelectual y que los coautores se forman los grupos sociales. Las Redes de coautoría nos proporcionan una base empírica para el análisis del cambio científico basado en las redes de cocitación o acoplamiento bibliográfico.

Referencias

Amin, M.; Mabe, M. Impact factors: Use and abuse. *Perspectives in Publishing*, Langford Lane, n.1, p.1-6, Oct. 2000. Available from: <http://cdn.elsevier.com/assets/pdf_file/0014/111425/Perspectives1.pdf>. Cited: Jul. 20, 2013.

Bourdieu, P. *Outline of a theory of practice*. New York: Cambridge University, 1977.

Bradford, S. C. Sources of information on specific subject. *Engineering: An illustrated weekly journal*, v.137, n. 3550, p.85-86, 1934.

Chen, C. Mapping scientific frontiers: *The quest for knowledge visualization*. New York: Springer-Verlag, 2003.

Cronin, B. *The citation process: The role and significance of citations in scientific communication*. London: Taylor Graham, 1984.

Egghe, L.; Rousseau, R. Transfer principles and a classification of concentration measures. *Journal of the American Society for Information Science*, v.42, n.7, p.479-489, 1991.

Frakes, W.B. Stemming Algorithms. In: Frakes, W.B.; Baeza-Yates, R. (Ed.). *Information retrieval: Data structures and algorithms*. Englewood Cliffs: Prentice-Hall, 1992. p.131-160.

Frame, J.D.; Narin, F.; Carpenter, J.P. The distribution of world science. *Social Studies of Science*, v.7, n.4, p.501-16, 1977.

Freeman, L.C. *The Development of social network analysis: A study in the sociology of science*. Vancouver: Empirical, 2004.

Garfield, E. Quantitative analysis of the scientific literature and its implications for science policymaking in Latin America and the Caribbean. *Bulletin of PAHO*, v.29, n.1, p.87-95, 1955.

Gibbs, W.W. Lost science in the third world. *Scientific American*, v.273, n.2, p.76-83, 1995.

Granovetter, M. Economic action and social structure: The problem of embed-dedness. *American Journal of Sociology*. v.91, n.3, p.481-510, 1985.

Gross, P.L.K.; Gross, E.M. College libraries and chemical education. *Science*, v.66, n.1713, p.358-389, 1927.

Heider, F. Attitudes and cognitive organization. *Journal of Psychology*, n.21, p.107-112, 1946.

Honkela, T. *et al.* Exploracion of full-text databases with self-organizing maps. In: International Conference on Neural Networks, 1., 1996, Piscataway. *Proceedings...* Piscataway: IEEE Service Center, 1996. p.56-61.

Kohonen, T. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, v.43, n.1, p.59-69, 1982.

Kucera, H.; Francis, N. *Computational analysis of presentday American English*. Providence, RD: Brown University Press, 1967.

Litwin, H. The network shifts of elderly immigrants: The case of Soviet Jews in Israel. *Journal of Cross Cultural Gerontology*, n.12, p.45-60, 1997.

Lotka, A.J. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, v.16, n.12, p.317-323, 1926.

- Moreno, J.L. *Who shall survive?* New York: Beacon House, 1934.
- Narin, F.; Hamilton, K.; Olivastro, D. The increasing linkage between U.S. technology and public science. *Research Policy*, v.26, n.3, p.317-330, 1997.
- Organization for Economic Cooperation and Development. *Manual de Frascati: propuesta de norma práctica para encuestas de investigación y desarrollo experimental*. Paris: OCDE, 1975.
- Price, D.J.S. *Little science, big science*. New York: Columbia University, 1964.
- Salton, G.; Buckley, C.; Yu, C.T. An evaluation of term dependence models in information retrieval. In: Annual ACM Conference on Research and Development in Information Retrieval, 5., 1982, Berlin. *Proceedings...* Berlin: Springer Heidelberg, 1982. p.151-173.
- Schubert, A.; Glänzel, W.; Braun, T. Scientometric datafiles: A comprehensive set of indicators on 2649 journals and 96 countries in all major science fields and subfields 1981-1985. *Scientometrics*, v.16, n.1, p.3-478, 1989.
- Thompson, J.W. The death of the scholarly monograph in the humanities? Citation patterns in literary scholarship. *Libri*, v.52, n.3, p.121-36, 2002.
- Thomson Reuters. *Products & Services*, 2010. Available from: <http://thomsonreuters.com/products_services/science/science_products/a-z/isi_web_of_knowledge/>. Cited: Jun. 24, 2013.
- White, H.D. Author-centered bibliometrics through CAMEOs: Characterizations automatically made and edited online. *Scientometrics*, v.51, n.3, p.607-637, 2001.
- White, H.D.; McCain, K.W. Bibliometrics. *Annual Review of Information Science and Technology*, n.24, p.119-186, 1989.
- Zipf, G.K. *Human behavior and the principle of least effort*. Addison-Wesley: Cambridge Mass, 1949.

Las prácticas de citación como interpretantes semióticos de acreditación de saberes locales en astronomía: México 1952-1972¹

Citation practices as semiotic interpretants of certification of local knowledge in astronomy: Mexico 1952-1972

Francisco COLLAZO-REYES²

Xochitl FLORES-VARGAS³

Mitzi Lizeth MUÑOZ-GARCÍA³

Miguel Ángel PÉREZ-ANGÓN²

Resumen

Se estudia la literatura científica en el área de astronomía publicada en México. Para ello se utilizan los textos publicados en el Boletín de los Observatorios de Tonantzintla y Tacubaya, durante un periodo de cambios importantes en la producción y publicación de conocimientos científicos en México, 1952-1972. Se busca identificar, en los textos científicos del Boletín, marcas de afiliación del texto local a los patrones de la comunicación científica moderna. Para ello revisamos las prácticas de citación seguidas en las publicaciones, particularmente las funciones de comunicación de las referencias como elementos retóricos de persuasión, en el proceso de formación del texto científico. Se utiliza una metodología de análisis semiótico basada en el modelo triádico de formación de conceptos. A través de esta metodología fue posible mostrar nuevos roles de los textos del Boletín en el sistema de la comunicación científica internacional. Interpretantes semióticos que dan cuenta de dos funciones de comunicación adquiridas por los textos: una como fuentes acreditadas y otra como fuentes acreditantes, en el proceso mutuo de certificación de conocimientos seguido en las revistas científicas.

Palabras clave: Análisis semiótico. Astronomía. Bibliografías citadas. Intertextualidad. México. Periódico científico.

Abstract

We studied the scientific literature in the field of astronomy published in the Mexican Journal, Boletín de los Observatorios de Tonantzintla y Tacubaya, from 1952 to 1972. We analyzed the articles published during this important period of great changes in the production and dissemination of scientific knowledge. Our goal is to identify affiliation practices in the patterns of scientific communication. The characteristics of citation practices were reviewed, particularly the communication functions of references as rhetorical elements of persuasion. Semiotic analysis was used, particularly the triadic model of concept formation. Accordingly, it was possible to demonstrate new roles of the Boletín articles within the framework of the system of international communication. Semiotic interpretants reveal two communication functions acquired by the texts: in the first case, sources are accredited, and the second case, sources are accrediting, in a mutual process of certification of knowledge found in scientific journals.

Keywords: Semiotic analysis. Astronomy. Citation bibliographies. Intertextuality. México. Scientific journals.

¹ Trabajo presentado en el VII Seminario Internacional sobre Estudios Cuantitativos y Cualitativos de la Ciencia y la Tecnología "Prof. Gilberto Sotolongo Aguilar" en XIII Congreso Internacional de Información - INFO' 2014. Habana, Cuba.

² Centro de Investigación y de Estudios Avanzados, Departamento de Física, Biblioteca de Ciencias Exactas. Av. Instituto Politécnico Nacional 2508, Colonia San Pedro Zacatenco, 07360, Ciudad de México, México. Correspondencia a nombre de/Correspondence to: F. COLLAZO-REYES. E-mail: <fcollazo@fis.cinvestav.mx>.

³ Centro de Investigación y de Estudios Avanzados, Biblioteca de Ciencias Exactas, Área de Bibliometría. Ciudad de México, México.

Recibido el día 10/6/2014 y aceptado para su publicación el 12/9/2014.

Introducción

La intertextualidad referencial (Porter, 1986; Haberer, 2007) forma parte de la naturaleza del discurso científico (Bazerman, 1988a, 1993). Es un componente bibliográfico que ha incrementado su importancia en el proceso de evolución del género científico moderno (Bazerman, 1988b) y lo ha distinguido de otros estilos literarios. Esta importancia se puede observar en la tendencia generalizada de incremento, en el tiempo, del número de referencias por trabajo, principalmente en las ciencias exactas, naturales y las ingenierías. Este hecho ha estado acompañado de distintas interpretaciones sobre las funciones de las referencias en el texto científico. La literatura en bibliometría/cienciometría da cuenta de este debate (Cronin, 1981; Cozzens, 1989; Bornmann & Daniel, 2008).

Los recursos argumentativos de estas propuestas han estado influenciados, históricamente, por dos concepciones distintas, ambas relacionadas con el desarrollo de la ciencia. Una influenciada por las teorías de Merton (1968). Este considera la ciencia como una institución normativa gobernada por un sistema interno de normas, recompensas y sanciones, que guían y constriñen las acciones de los científicos (Cronin, 1984; White, 2004). Las citas, como parte de este sistema, son consideradas como eventos de reconocimiento o recompensa. Existe una abundante literatura y consenso sobre el concepto de la cita como una relación de reconocimiento entre el documento que cita (citante) y los documentos citados. A partir de esta interpretación se han desarrollado las concepciones dominantes sobre las funciones de las citas, como es el caso del factor de impacto en todas las modalidades (Cronin, 1984).

Otras propuestas conciben la ciencia como prácticas socialmente construidas (Knorr-Cetina, 1981; Latour, 1987). Aquí el uso de las referencias está gobernado por una función de persuasión (Gilbert, 1977; Paul, 2000; Moed, 2005). Como componentes del arte retórico de la persuasión (Restrepo Forero, 2004; Carmona Sandoval, 2013). Bajo esta concepción se asume que una n. 14 de las mayores motivaciones de citar está conectada con el deseo de persuadir lectores (MacRoberts, M.H. & MacRoberts, B.R., 1987). En esta interpretación de las referencias como dispositivos retóricos para persuadir (Restrepo Forero, 2004; Carmona Sandoval, 2013), el autor

tiende a citar trabajos importantes o autorizados, de tal manera que el trabajo que los cita participa de su autoridad (Gilbert, 1977). Esta concepción de las referencias ha permanecido un poco olvidada (White, 2004; Davis, 2009). Sin embargo, el concepto de persuasión resulta más adecuado para los trabajos de corte histórico, como el que aquí realizamos. Creemos que los cambios en la formación de los patrones de citación, como parte del proceso histórico de formación del texto científico, forman parte de un escenario mayor. La creación de condiciones que hacen posible la producción de un sentido (Beuchot, 2004) de filiación de la ciencia local a los cánones de la ciencia internacional. También, que es posible caracterizar este sentido de filiación, a través del estudio de las relaciones entre textos (intertextualidad referencial), como signos de acreditación, entre textos, que en la medida que se repiten, significan la relación entre discursos con contenidos afines.

En los procesos de conformación de los sistemas de comunicación científica locales, basados en las fuentes e idioma doméstico, las prácticas de citación moderna tardaron en formar parte de la estructura de los contenidos y el estilo argumentativo de los textos científicos (Luna-Morales *et al.*, 2009). En México, los *Anuarios de los Observatorios de Tacubaya y de Chapultepec*, *Las Memorias de la Sociedad Científica Antonio Alzate*, y *El Boletín de la Sociedad Astronómica de México*, cubren el periodo histórico más largo de conocimientos publicados en astronomía, 1881-1950 (González-Tapia & Muñoz-García, 2013). El cuerpo de literatura acumulado en estas revistas contiene los temas, tipos de investigación, estilos, idioma, métodos e instrumental científico, utilizados en este tipo de investigación observacional. Este periodo, está dominado por temáticas relacionadas con la astronomía clásica o astronomía de posición y las observaciones relacionadas con la elaboración de la Carta y el Catálogo astrofotográfico del cielo de México, como parte de un proyecto internacional (Bok, 1983; Pismis, 1987; Bartolucci, 1991, 2000).

Estas prácticas de investigación dieron lugar a las estructuras de un sistema de publicación y comunicación científica local. Dependiente de los criterios mínimos seguidos para publicar en las fuentes mencionadas.

Predominó un estilo de contenidos con funciones de comunicación orientadas a la divulgación de conocimientos astronómicos generales sobre diversos aspectos: cometas, eclipses, luna, cielo, sol, planetas (Júpiter, Marte, Saturno), estrellas, observatorios, constelaciones, entre otros; así como desarrollo de trabajos sobre aspectos aplicados: determinar posiciones geográficas específicas, elaborar mapas e informes periódicos sobre datos astronómicos reportados por mes, así como reportes sobre fenómenos astronómicos importantes (González-Tapia & Muñoz-García, 2013). En el desarrollo de estos contenidos muy pocas veces se recurría a las referencias y éstas no forman parte sustantiva del estilo argumentativo de los textos. Los primeros trabajos publicados del área en revistas internacionales (*Transactions-American Geophysical Union* y *Astronomical Journal*), tienen estas mismas características: tratan aspectos descriptivos referentes al trabajo administrativo y al equipo utilizado en los observatorios, no incluyen referencias (Gallo, 1939; Rivera-Terrazas, 1950). Los primeros antecedentes de prácticas modernas de citación, en la astronomía generada en México, se encuentran en un trabajo escrito por Munch (1947) con adscripción al Observatorio de Tacubaya. Es un trabajo teórico que incluye 8 referencias y ha recibido 19 citas bibliográficas.

Históricamente la vía común de acreditación internacional del conocimiento generado localmente ha sido a través de la publicación arbitrada en revistas reconocidas en la corriente principal de cada área. Aquí nos referimos a prácticas científicas desarrolladas y difundidas en las condiciones adversas de un contexto local-local (Cueto, 1997). Se trata de los conocimientos generados en observatorios locales y publicados en una revista local, el Boletín de los Observatorios de Tonantzintla y Tacubaya (BOTT). Una revista que en su momento de vigencia, 1952-1972, estuvo relegada del *Science Citation Index* (SCI), aunque este índice da cuenta de las citas recibidas, por trabajos del BOTT, de revistas cubiertas en este índice. Con el tiempo ha adquirido mayor visibilidad internacional aún después que dejó de publicarse. Actualmente la colección completa de textos se encuentra disponible en línea en la Universidad Nacional Autónoma de México (UNAM) <<http://www.astroscu.unam.mx/bott/>> y en el sistema *Astrophysic Data System* (ADS) / *National Aeronautics and*

Space Administration (NASA) <http://adsabs.harvard.edu/bib_abs.html>. En el año internacional de la astronomía, 2009, la comunidad científica mexicana en astronomía organizó un congreso internacional "Highlights of Boletín de los Observatorios de Tonantzintla y Tacubaya" (Torres-Peimbert *et al.*, 2011). Tuvo como finalidad revisar el rol del BOTT en el desarrollo de la astronomía moderna en México. Los trabajos presentados ofrecen una revisión de la producción histórica del Boletín, principalmente los trabajos más citados. Se abordan aspectos referentes a los resultados, las contribuciones, relevancia, temas iniciales, circunstancias del contexto, instrumental científico y técnicas utilizadas; a la luz de los más recientes desarrollos en el campo (López-Cruz, 2011; Cornejo-Rodríguez, 2011; Abt, 2011a, 2011b).

En este trabajo, identificamos y revisamos las estructuras de relaciones de comunicación científica, desarrolladas en las prácticas de investigación y publicación de los trabajos del BOTT. Para ello utilizamos las referencias hechas y las citas recibidas en los textos. Lo anterior con el fin de caracterizar estos flujos de comunicación, como un método de aproximación al proceso de acreditación de los textos del Boletín en los distintos discursos a que pertenecen las referencias hechas y los discursos de donde proceden las citas recibidas, en los textos analizados del área de astronomía.

Los artículos científicos como nodos de una red, establecen relaciones con otros artículos a través de las referencias y las citas. En este ámbito, las referencias del artículo pueden desarrollar, desde el punto de vista semiótico, distintos interpretantes. En este caso nos referimos a dos: (1) las referencias como elementos de acreditación de los textos citantes, y (2) las referencias, reasignadas como citas, como elementos acreditantes de los textos citados. Para identificarlos utilizamos el análisis triádico de los conceptos de la semiótica de Peirce (1992), Amozorrutia (2010) e Vidales-González, (2010). Se utilizan dos esquemas triádicos: el primero se refiere al rol de la referencia como parte del artículo y el segundo a la referencia reasignada como cita Small (1978), Cronin (1981, 2000) e Collazo-Reyes (2012). El primer interpretante se desarrolla al momento de incluir referencias en una publicación, hechas a otros trabajos previamente publicados y, el segundo, cuando un trabajo aparece citado, después de su publicación, en otros trabajos. En

ambos casos se identifican las direcciones de los flujos de comunicación desarrollados entre textos. En el caso de las referencias, identificamos los flujos que van hacia la literatura referida en tres contextos distintos: local, regional e internacional. En el caso de las citas los flujos también pueden proceder de contextos locales, regionales o internacionales.

Los resultados muestran que los contenidos de los trabajos publicados en el BOTT, forman parte de la red de relaciones internacionales de intertextualidad referencial entre textos (Durañona *et al.*, 2006) aceptados en revistas científicas y que, en este escenario, juegan roles de textos acreditados y de textos que acreditan, en un proceso mutuo de certificación de conocimientos seguido en el sistema de comunicación científica moderna. Estos aspectos son eventos emergentes, en el proceso histórico de formación del texto científico moderno en astronomía y le aportan novedades cualitativas a la estructura de los contenidos, las funciones de comunicación y el estilo del género científico en México.

Métodos

Se utilizó como fuente principal de análisis 154 textos completos de los trabajos publicados en el BOTT; las referencias hechas en el cuerpo de los trabajos, durante el periodo 1952-1972, y las citas recibidas en el periodo 1972-2013. Las referencias se recuperaron a través de una búsqueda en los contenidos de todos los textos. En el caso de las referencias incompletas, se consideraron las que incluían como información mínima el nombre del autor o la fuente y el año de publicación. Las citas fueron recuperadas del *Web of Science* (WoS), a través de la modalidad de búsqueda *Cited Reference Search*, por cada uno de los nombres de los autores participantes en los trabajos publicados en el BOTT. Se trata de citas hechas en revistas incluidas en el WoS para trabajos en el BOTT, que es una revista sin cobertura en este sistema de información. Las referencias y las citas fueron organizadas en archivos relacionales independientes, con una estructura de datos que incluye cinco campos de información: autor, año, fuente, volumen y página. Se desarrolló una base de datos relacional que incluye los 154 trabajos publicados en el BOTT, asociados a través

de tablas relacionales, con los archivos de las referencias bibliográficas hechas en los trabajos y las citas recibidas.

Modelo de análisis semiótico de las referencias y las citas

El modelo semiótico de análisis tradicional consiste de tres elementos: signo o representamen (primeridad), objeto (segundidad) e interpretante (terceridad). El signo es la referencia bibliográfica que sustituye a un objeto físico que puede ser el registro bibliográfico, un artículo, capítulo, libro, o memoria. El interpretante tiene que ver con un concepto emergente de acreditación internacional de los textos en astronomía. Una novedad cualitativa en sus contenidos desarrollada a través de funciones de comunicación propias de las referencias y las citas. El desarrollo de este sentido de acreditación de los textos se explica a través de los modelos tríadicos de las Figuras 2 y 3.

Resultados

Encontramos un total de 1.139 referencias publicadas en 81 fuentes de distintos tipos. El 78,0% corresponden a artículos en revistas incluidas en el WoS, el 12,0% a trabajos en fuentes no publicadas como conferencias, tesis, reportes, entre otros; el 6.5% se encuentran en fuentes publicadas pero que no son

La referencias interpretadas como signo

Análisis tríadico de C. Peirce

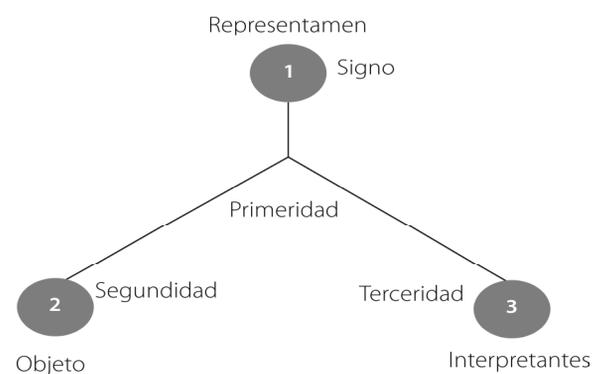


Figura 1. Modelo semiótico de análisis tríadico de C. Peirce.

Fuente: Elaboración propia, 2014.

cubiertas por el WoS. El 3.5% restante se trata de referencias sin información sobre la fuente de publicación.

De los 154 trabajos publicados en el BOTT, 129 recibieron 2,405 citas en el periodo 1961-2013. Las citas proceden de trabajos publicados en 116 revistas, publicadas en distintos países, con una alta concentración de más del 80% en revistas de United States of America. Cerca del 6% proceden de trabajos publicados en revistas (*Revista Mexicana de Astronomía y Astrofísica*) y memorias de congresos celebrados en México.

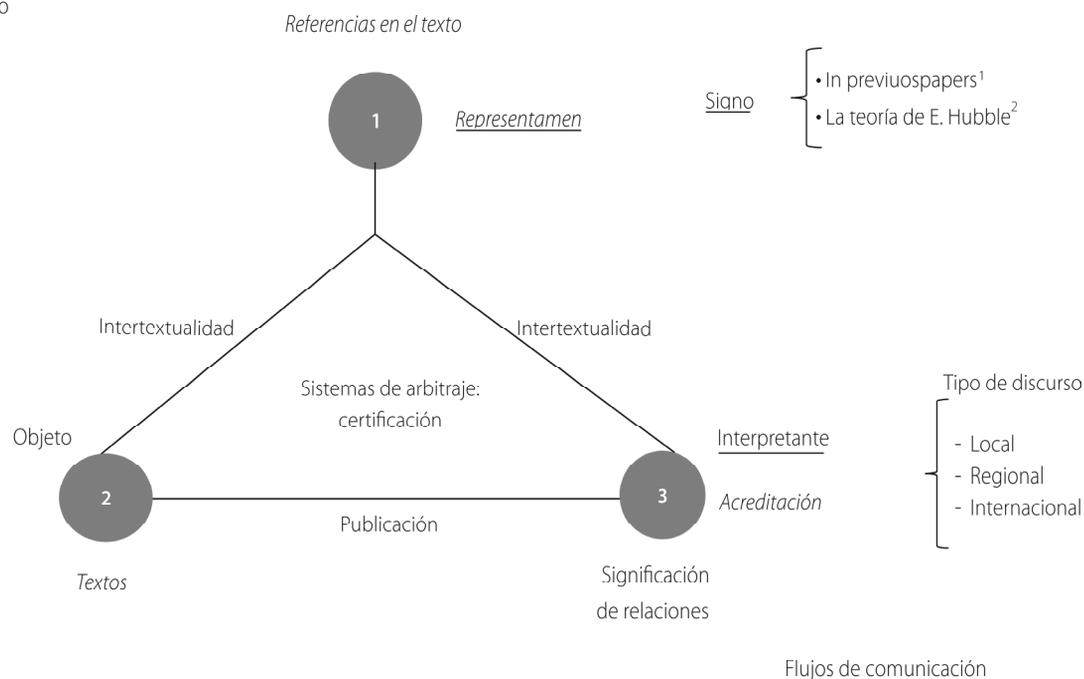
Análisis semiótico de referencias

En primer lugar, se analiza el proceso de significación de la referencia como un elemento de

persuasión en la retórica del discurso científico (Restrepo Forero, 2004). Durante las distintas prácticas de construcción, arbitraje y publicación de los textos, las referencias desarrollan un interpretante de acreditación del estilo argumentativo de los contenidos. El nivel de primeridad (representamen), se refiere a la etapa de construcción de los textos. Cada vez que se utiliza una referencia bibliográfica se crea una relación de intertextualidad con otro texto, generalmente, ya publicado. En esta etapa el autor utiliza elementos de la referencia bibliográfica o de los textos citados, como convenciones de redacción, para representar el texto citado. Las prácticas de intertextualidad referencial utilizadas en el BOTT tienen las siguientes características: se utilizó, como signo principal de las referencias, números consecutivos dentro del texto en forma de super índices. Cada número apunta a una referencia

Las referencias como interpretantes semióticos de persuasión

Análisis trádico



Referencias

1. W. J. Luyten. Ap. j. **58**, 75, 1953. A. J. **61**, 261, 1956.
2. Ap. J. vol 113, No. 3 November 1951.

Figura 2. Modelo de análisis trádico de la referencia como interpretante semiótico de acreditación.

Fuente: Elaboración propia, 2014.

bibliográfica de la obra citada, consignadas al final del texto en la sección de referencias. Para las notas de pie de página se utilizó como signo el asterisco en el texto y consignan aspectos diferentes a las referencias bibliográficas. En esta primera instancia del modelo, las referencias se reasignan como parte del contenido de los trabajos citantes, como elementos de apoyo al estilo argumentativo y de persuasión propios del género científico. Las referencias más comunes son para artículos en revistas internacionales, con un promedio de 7 referencias cada uno.

El segundo nivel (objetos) de la triada, se refiere a las referencias bibliográficas completas, organizadas como una unidad específica en la división de contenidos de los textos físicos, como objetos completos. Una sección que incluye un promedio de 7 referencias por documento, hechas a distintos tipos de fuentes que dividimos en cuatro tipos: (1) para artículos de revistas, no se incluye el título del trabajo, solamente el nombre del autor, título de la revista, volumen, página y año; (2) en el caso de libros, capítulos, tesis y congresos, se incluye el autor, título de la obra, lugar de edición, editorial y año; (3) cuando se hace referencia a los nombres de los autores del artículo, la referencia solo incluye el título de la revista, volumen, página y año; (4) en las referencias a las conversaciones privadas, datos no publicados y la correspondencia personal, se incluye el nombre del autor. Cada referencia puede desarrollar un rol distinto como fuente de acreditación del texto citante, de acuerdo con las siguientes características. El nivel de comunicación científica de pertenencia: local, regional o internacional; el tipo de investigación, teórica, experimental, difusión; el idioma de publicación; el prestigio de los autores en el campo de investigación; el tipo de revista; con o sin arbitraje, idioma, país de edición, y presencia en índices internacionales; actualidad de la información, identificada a partir del año de publicación. Estas características, actúan en favor o en contra de los textos escritos en los procesos de arbitraje y negociación de su publicación.

En el tercer nivel del modelo, los procesos de acreditación de los textos como objetos hechos públicos en revistas científicas, son eventos de comunicación que se repiten y generan circunstancias de producción de sentido. En este escenario de la comunicación y de acuerdo con el modelo de la Figura 1, los elementos que

conforman el contenido de los textos del BOTT pueden desarrollar distintos interpretantes semióticos. En el caso que nos ocupa, cada referencia, en el texto, establece relaciones de intertextualidad con la literatura de algún tipo de discurso. Las relaciones pueden apuntar hacia la literatura y discursos considerados como periféricos, regionales o internacionales. En la medida que estas relaciones se repiten, con trabajos de un mismo discurso, las relaciones se distinguen y significan como flujos de comunicación entre textos con discursos afines. Estas relaciones se interpretan como eventos de acreditación del discurso de los documentos publicados en el BOTT en el discurso de los documentos referidos. En términos semióticos, este proceso de acreditación es un interpretante que desarrollan las referencias, como una función comunicativa que caracteriza su rol en el sistema de comunicación científica.

De este proceso de acreditación surgen distintos flujos de comunicación, mostrados en la Tabla 1 (A - Referencias hechas). Cuando estas relaciones apuntan hacia la literatura del mismo país o de países de la misma región, tenemos acreditaciones de tipo endogámica y regional, respectivamente. Cuando se refieren a la literatura internacional tenemos acreditaciones en el discurso de corriente principal en el área de astronomía.

Análisis semiótico de las citas

En una etapa posterior a su publicación, los textos desarrollan un segundo interpretante de acreditación a través de las funciones de comunicación de las citas. Este sentido de las citas se construye a través de la misma relación tríadica de signo, objeto e interpretante. De acuerdo con el modelo de la Figura 3, en el primer nivel de la tríada, utilizamos la referencia no como parte del texto citante (según el modelo de la Figura 2). Nos interesa la referencia utilizada como signo del texto físico citado, representado por el objeto del segundo nivel de la tríada. Cuando un autor utiliza una referencia bibliográfica en alguna parte del contenido de su trabajo, las referencias se reasignan como citas y se establece una relación de comunicación, donde el documento que hace la cita es instanciado como documento citante y el que la recibe como documento citado. Cada vez que el objeto citado (segundo nivel de la tríada), es referido en

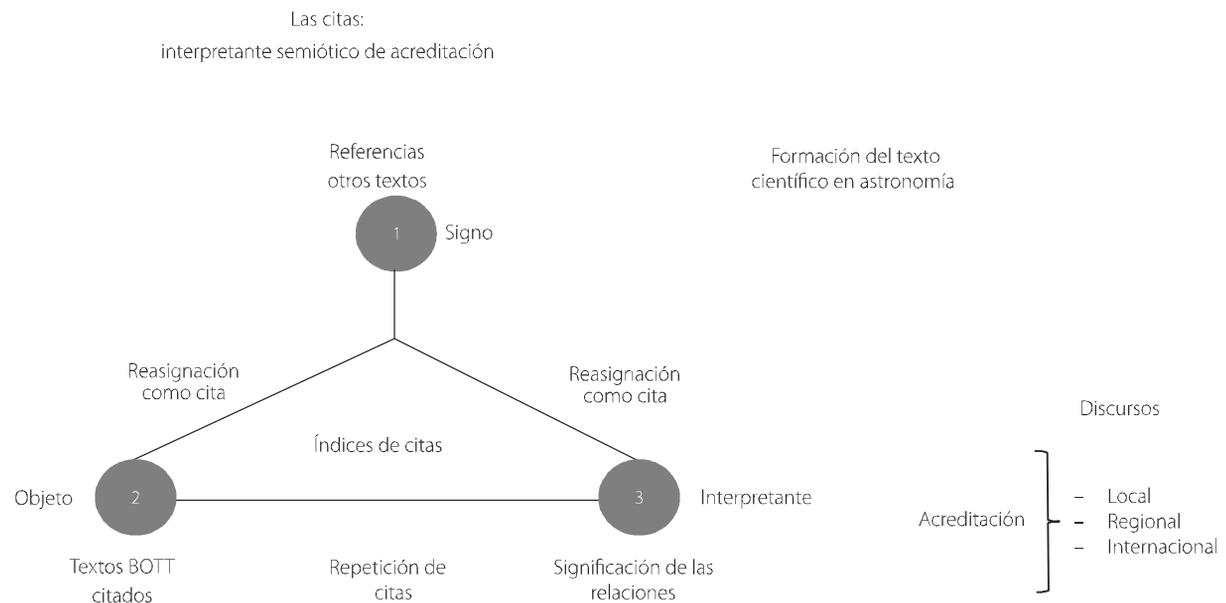


Figura 3. Análisis semiótico de la cita bibliográfica como elemento acreditante.

Fuente: Elaboración propia, 2014.

Tabla 1. Flujos de comunicación científica resultantes de los trabajos publicados en el Boletín de los Observatorios de Tonantzintla y Tacubaya: A) referencias hechas, 1952-1972; B) citas recibidas, 1961-2013.

A) Referencias hechas						
Flujos	Autores Citantes	Revista	Revista Citada	Relaciones	%	Flujos de comunicación
1	L	BOTT	E	885	77.7	Local-local-externos
2	L	BOTT	L	69	6.06	Endógenos
3	E	BOTT	E	158	13.87	Externo-local-externo
4	E	BOTT	L	18	1.58	Externo-local-local
5	R	BOTT	E	6	0.53	Regional-local-externo
6	R	BOTT	L	3	0.26	Regional-local-local

B) Citas recibidas						
Flujos	Autore Citante	Revista	Revista Citada	Relaciones	%	Flujos de comunicación
1	L	L	L	109	4.53	Endógenos
2	L	E	L	234	9.73	Local-externo-local
3	E	E	L	1998	83	Externo-externo-local
4	E	L	L	46	1.91	Externo-local-local
5	R	E	L	18	0.75	Regional-externo-local

Nota: L = local; E = externo; R = regional.

Fuente: Elaboración propia, 2014.

algún tipo de literatura publicada, se considera como un signo indicativo de acreditación del objeto citado en el discurso de la literatura del objeto citante. De acuerdo con el tercer nivel de la tríada, los textos del BOTT, desarrollan un interpretante de textos acreditantes en el sistema de comunicación científica, cada vez que son

citados por otros textos. A diferencia de las referencias, aquí las citas se interpretan como eventos acreditantes, donde los textos del BOTT funcionan como elementos de apoyo a la retórica argumentativa y de persuasión de los textos citantes. Así, en la medida que la procedencia de las citas se repiten, con trabajos pertenecientes a un

mismo discurso (local, regional, internacional), las relaciones se distinguen y significan como flujos de comunicación entre textos de discursos afines. Estos flujos traducidos a estructuras de comunicación, se utilizan como testimonios del proceso que certifica el discurso de los documentos citados en el discurso de los documentos citantes. El interpretante semiótico de las citas acredita el discurso de los documentos citados en distintos tipos de literatura que se muestran en la Tabla 1.

La Tabla 1 (B - Citas recibidas), muestra cinco flujos de comunicación resultantes de las citas hechas a los trabajos del BOTT. Incluye flujos endógenos, local-externos y externos-locales, principalmente.

L = local; E = externo; R = regional

La Tabla 1 (A) referencias hechas, muestra seis flujos resultantes de las referencias hechas en los trabajos del BOTT. Un poco más de tres cuartas partes (77.7%) de las referencias fueron hechas por autores adscritos a instituciones locales para trabajos publicados en revistas extranjeras. La Tabla 1 (B), muestra los títulos de las revistas citadas y el número de referencias. Cerca de la mitad de las referencias fueron para trabajos publicados en *Astrophysical Journal*. Los autores adscritos a instituciones externas que publicaron en el BOTT realizaron cerca del 15% de las referencias. Las referencias hechas por autores locales a trabajos publicados en fuentes locales, principalmente el BOTT, representan el 6% y se denominaron como flujos endógenos. Los autores regionales, adscritos a instituciones de países de América Latina, realizaron menos de 10 citas.

De acuerdo con la Tabla 1 (B Citas recibidas), un poco más de cuatro quintas partes (83%) de las citas fueron hechas por autores adscritos a instituciones extranjeras en trabajos publicados en fuentes internacionales. El segundo flujo más importante de citas, procede de autores locales que publican en revistas externas. También existen flujos endogámicos de acreditación entre autores, trabajos y fuentes locales. Los autores regionales, adscritos a instituciones de países de América Latina, realizaron menos de 18 citas en revistas externas.

De acuerdo con estos flujos procedentes, principalmente, de autores y revistas extranjeras, los trabajos del BOTT están acreditados en el mismo sistema

de información de los trabajos y las revistas citantes, que es el del discurso dominante en astronomía a nivel internacional.

Revistas citadas y citantes

De acuerdo con las fuentes citadas de la Tabla 2 (A), cerca de la mitad de las referencias fueron hechas a trabajos publicados en *Astrophysical Journal*, una de las revistas de mayor prestigio en el campo. El 10% de las referencias fueron para trabajos publicados en el propio BOTT que aparece en segundo lugar. Las referencias para libros y tesis suman cerca de un 4%. Las fuentes de publicación citadas son editadas en distintos países (USA, México, Gran Bretaña, Japón, Holanda, Francia, Alemania), sin embargo, más de las tres cuartas partes de las referencias son para trabajos publicados en fuentes editadas en USA. De acuerdo con estos flujos repetitivos de intertextualidad referencial, los trabajos del BOTT están acreditados, principalmente, en trabajos y revistas del discurso dominante a nivel internacional.

La Tabla 2 (B fuentes citantes), muestra los títulos de las revistas citantes y el número de citas. El 85% de las citas están hechas en los primeros 11 títulos de revistas. Nueve de estas, están publicadas en USA. Se destacan los dos primeros títulos *Astrophysical Journal* y *Astronomy & Astrophysics*, que suman más de mil citas, cerca del 50% del total. El 12% de las citas proceden de trabajos publicados en una revista de la Gran Bretaña (*Monthly Notices of the Royal Astronomical Society*), ubicada en tercer lugar. Las revistas locales, lugares 7 (Revista Mexicana de Astronomía y Astrofísica) y 9 (Revista Mexicana de Astronomía y Astrofísica, Serie de Conferencias), realizaron el 5% de las citas. Los títulos de las revistas ubicadas en los lugares 1, 3 y 4 de las Tablas 2 (A) y 2 (B), respectivamente, son las mismas. Esto es señal de una importante correlación entre las revistas que acreditan los trabajos del BOTT, tanto a través de las referencias como a través de las citas.

Discusión y Conclusión

Las prácticas de investigación científica desarrolladas en los laboratorios de Tonantzintla y Tacubaya, generaron un espacio de encuentro entre componentes de una práctica científica: especialistas en

Tabla 2. Intertextualidad referencial resultante de los trabajos publicados en el Boletín de los Observatorios de Tonantzintla y Tacubaya: A) Fuentes citadas, 1952-1972; B) fuentes citantes, 1961-2013.

A) Fuentes citadas, 1952-1972		
Número	Revistas	Total
1	<i>Astrophysical Journal</i>	525
2	Boletín de los Observatorios de Tonantzintla y Tacubaya	89
3	<i>Monthly Notices of the Royal Astronomical Society</i>	58
4	<i>Astronomical Journal</i>	51
5	Book	48
6	<i>Astrophysica Journal Supplement Series</i>	33
7	<i>Publications of the Astronomical Society of Japan</i>	27
8	<i>Bulletin of the Astronomical Institutes of the Netherlands</i>	24
9	<i>Publications of the Astronomical Society of the Pacific</i>	20
10	<i>International Astronomical Union. Symposium</i>	17
11	QB	15
12	<i>Annual Review of Astronomy and Astrophysics</i>	14
13	Ph.D. Thesis	12
14	<i>Lowell Observatory Bulletin</i>	12
15	<i>Annales d'Astrophysique</i>	11
16	<i>Zeitschrift für Astrophysik</i>	10
17	Otros	173
Total		1.139
B) Fuentes citadas, 1961-2013		
Número	Revistas	Total
1	<i>Astrophysical Journal</i>	719
2	<i>Astronomy & Astrophysics</i>	369
3	<i>Monthly Notices of the Royal Astronomical Society</i>	290
4	<i>Astronomical Journal</i>	178
5	<i>Publications of the Astronomical Society of the Pacific</i>	130
6	<i>Astrophysical Journal Supplement Series</i>	98
7	Revista Mexicana de Astronomía y Astrofísica	86
8	<i>Astrophysics and Space Science</i>	60
9	Revista Mexicana de Astronomía y Astrofísica-Serie de Conferencias	72
10	<i>Annual Review of Astronomy and Astrophysics</i>	37
11	<i>Astronomy & Astrophysics Supplement Series</i>	30
12	<i>Astronomicheskii Zhurnal</i>	27
13	Observatory	26
14	<i>Soviet Astronomy Aj Ussr</i>	24
15	<i>Publications of the Astronomical Society of Japan</i>	18
16	<i>Astronomische Nachrichten</i>	14
17	Otros	227
Total		2.405

Fuente: Elaboración propia, 2014.

el área de astronomía, metodologías, prácticas de investigación, instrumental científico, conocimientos y dominio de los paradigmas científicos, y una fuente de difusión de los resultados de investigación. La coincidencia de estos elementos, se interpretan como aspectos determinantes, en la medida que incidieron en

el desarrollo de circunstancias estables que hicieron posible la construcción de prácticas científicas productivas en el área de astronomía. Esta condición de estabilidad, permitió la continuidad en la producción de resultados de investigación y la formación de estructuras de comunicación científica moderna.

El análisis trádico de la semiótica, como proceso de significación de sentido, es un recurso de utilidad para identificar distintos interpretantes que desarrollan las referencias y las citas, como funciones de comunicación en los textos. Estos interpretantes ayudan a caracterizar aspectos emergentes en el proceso de formación que sigue el género científico en cualquier país; como parte de una dinámica que acompaña la conformación de las estructuras de los sistemas de comunicación científica en formación, a cualquier nivel, local, regional o internacional.

Los resultados muestran una importante correlación entre los títulos de las revistas citadas y citantes, donde el punto de encuentro son los textos del BOTT. Esto significa que los contenidos de los trabajos publicados en el Boletín, forman parte de la red de relaciones internacionales de intertextualidad referencial entre textos aceptados en las revistas científicas más importantes en el área de astronomía: *Astrophysical Journal*; *Astronomy & Astrophysics*; *Monthly Notices of the Royal Astronomical Society*; *Astronomical Journal*; *Publications of the Astronomical Society of the Pacific*, entre otras. En este escenario de medios de difusión, donde se

construye el discurso dominante del área, las publicaciones del Boletín juegan roles de textos acreditados y que acreditan, o acreditantes; formando parte de un proceso mutuo de certificación de conocimientos seguido en el sistema de comunicación científica moderna. Estos aspectos, son eventos emergentes en el proceso histórico de formación del texto científico moderno en México. Aportan novedades cualitativas, al estilo argumentativo del género científico en el área de astronomía. Las novedades incluyen marcas del contexto local e internacional que trascienden los patrones endogámicos del sistema de comunicación científica local, dependiente de mecanismos endogámicos de acreditación de conocimientos y, tradicionalmente, basado en fuentes de difusión doméstica, idioma local y un estilo argumentativo escaso de recursos de intertextualidad referencial.

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología - Red de Complejidad, Ciencia y Sociedad.

Referencias

Abt, H.A. Symposium summary. *Revista Mexicana de Astronomía y Astrofísica*, v.39, p.117-118, 2011a.

Abt, H.A. The Tonantzintla search for high luminosity stars. *Revista Mexicana de Astronomía y Astrofísica*, v.39, p.65-68, 2011b.

Amozorrutia, J.A. De la semiótica a la cibernética: una estrategia para modelar sistémicamente el sentido. *Razón y Palabra*, v.15, n.72, p.1-24, 2010.

Bartolucci, J. Formación tardía de comunidades científicas: el caso de los astrónomos mexicanos. *Quiipu*, v.8, n.3, p.361-377, 1991.

Bartolucci, J. *La modernización de la ciencia en México: el caso de los astrónomos*. México: UNAM, 2000.

Bazerman, C. Making references: Empirical contexts, choices, and constraints in the literary creation of the Compton effect. In: Bazerman, C. *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison: University of Wisconsin Press, 1988a. p.187-234.

Bazerman, C. *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison: University of Wisconsin Press, 1988b, 356p.

Bazerman, C. Intertextual self-fashioning: Gould and Lewontin's representations of the literature. In: Selzer, J. (Ed.). *Understanding scientific prose*. Madison: University of Wisconsin Press, 1993. p.20-41.

Beuchot, M. *La semiótica: teoría del signo y el lenguaje en la historia*. México: Fondo de Cultura Económica, 2004.

Bok, B.J. Mexican astronomy, 1930-1950. *Revista Mexicana de Astronomía y Astrofísica*, v.7, (Esp.), p.21-25, 1983.

Bornmann, L.; Daniel, H.D. What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, v.64, n.1, p.45-80, 2008.

Carmona Sandoval, J.C. Discurso y artículo científico: una aproximación retórica. *Raximhai*, v.9, n.1, p.117-152, 2013.

Collazo-Reyes, F. Modelo teórico metodológico para analizar la visibilidad internacional de las ciencias físicas en América Latina y el Caribe. 2012. Tesis (Doctorado en Bibliotecología y Estudios de la Información). Universidad Nacional Autónoma de México, México, 2012.

Cornejo-Rodríguez, A. The first years of optics in México and the role of Boletín de los Observatorios de Tonantzintla y Tacubaya on its development. *Revista Mexicana de Astronomía y Astrofísica*, v.39, p.109-115, 2011.

- Cozzens, S.E. What does citations count? The rhetoric-first model. *Scientometric*, v.15, n.5, p.1437-447, 1989.
- Cronin, B. The need for a theory of citation. *Journal of Documentation*, v.37, n.1, p.16-24, 1981.
- Cronin, B. *The citation process: The role and significance of citation in Scientific Communication*. London: Taylor Graham, 1984.
- Cronin, B. Semiotic and evaluative bibliometrics. *Journal of Documentation*, v.56, n.4, p.440-453, 2000.
- Cueto, M. Science under adversity: Latin American Medical Research and American Private Philanthropy, 1920-1960. *Minerva*, v.35, n.3, p.233-245, 1997.
- Davis, P.M. Reward or persuasion? The battle to define the meaning of a citation. *Learned Publishing*, v.22, n.1, p.5-11, 2009.
- Durañona, M.A. et al. *Textos que dialogan: la intertextualidad como recurso didáctico*. Madrid: Comunidad de Madrid, 2006.
- Gallo, J. Magnetic work of the National Astronomical Observatory of the National University of México. *Transactions American Geophysical Union*, v.20, n.3, p.394, 1939.
- Gilbert, G.N. Referencing as persuasion. *Social Studies of Science*, v.7, p.113-122, 1977.
- González-Tapia, F.B.; Muñoz-García, M.L. Estudio histórico-bibliométrico de la literatura científica generada en astronomía en México, durante el siglo XX. 2013. Tesis (Licenciatura en Biblioteconomía). Escuela Nacional de Biblioteconomía y Archivonomía, Ciudad de México, México, 2013.
- Haberer, A. Intertextuality in theory and practice. *Literature*, v.49, n.5, p.54-67, 2007.
- Knorr-Cetina, K. *The manufacture of knowledge: An essay on the constructivist and contextual nature of science*. Oxford: Pergamon Press, 1981.
- Latour, B. *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press, 1987.
- López-Cruz, O. Informal introduction. *Revista Mexicana de Astronomía y Astrofísica*, v.39, p.xiii-xiv, 2011.
- Luna-Morales, M.E. et al. Early patterns of scientific production by mexican researchers in mainstream journals, 1900-1950. *Journal of the American Society Information Science and Technology*, v.60, n.7, p.1337-1348, 2009.
- MacRoberts, M.H.; MacRoberts, B.R. Testing the Ortega hypothesis: Facts and artifacts. *Scientometrics*, v.12, n.5-6, p.293-295, 1987.
- Merton, R.K. The Matthew effect in science. *Science*, v.159, n.5, p.56-63, 1968.
- Moed, H.F. What do references and citations measure? In: Moed, H.F. (Ed.). *Citation analysis research evaluation*. Netherlands: Springer, 2005. p.193-208.
- Munch, G. Model Solar Atmospheres. *Astronomical Journal*, v.106, n.2, p.217-223, 1947.
- Paul, D. In citing chaos: A study of the rhetorical use of citations. *Journal on Business Technology Communication*, v.4, n.2, p.185-216, 2000.
- Peirce, C.S. *The essential Peirce: Selected philosophical writings (1867-1983)*. Bloomington: Indiana University Press, 1992.
- Pismis, P. La astronomía en México. Hacia su etapa actual. *Revista Mexicana de Astronomía y Astrofísica*, v.14, n.1, p.35-42, 1987.
- Porter, J. E. Intertextuality and the discourse community. *Rhetoric Review*, v.5, n.1, p.35-47, 1986.
- Restrepo Forero, O. Retórica de la ciencia sin "retórica": sobre autores, comunidades y contextos. *Revista Colombiana de Sociología*, v.23, p.251-268, 2004.
- Rivera-Terrazas, M. The program of the Tonantzintla Schmidt camera. *Astronomical Journal*, v.55, n.3, p.65, 1950.
- Small, H.G. Cited documents as concept symbols. *Social Studies of Science*, v.8, n.3, p.327-340, 1978.
- Torres-Peimbert, S.; López-Cruz, O.; Peimbert, M. Historical note. *Revista Mexicana de Astronomía y Astrofísica*, v.39, p.vi-vii, 2011.
- Vidales-González, C. *Semiótica y teoría de la comunicación*. Tomo I. México: CECyTE, 2010, 135p.
- White, H.D. Reward, persuasión, and the Sokal Hoax: A study in citation identities. *Scientometrics*, v.60, n.1, p.93-120, 2004.

Producción científica cubana en Medicina y Salud Pública: Scopus 2003-2011¹

Cuban scientific production in Medicine and Public Health: Scopus 2003-2011

Grisel ZACCA-GONZÁLEZ²

Benjamín VARGAS-QUESADA³

Zaida CHINCHILLA-RODRÍGUEZ⁴

Félix de MOYA-ANEGÓN⁴

Resumen

El objetivo de este trabajo fue analizar la evolución del volumen y la visibilidad de la producción científica cubana en Salud Pública y en Medicina para determinar si siguen los mismos patrones de comunicación, y recomendar buenas prácticas de publicación. Se aplicaron indicadores bibliométricos de volumen, visibilidad y colaboración extraídos del portal *SCImago Institutions Rankings* a partir de datos de Scopus, para el área temática Medicina y la categoría *Public Health, Environmental and Occupational Health*, período 2003-2011. Cuba tiene una posición relativamente alta en los rankings de volumen de producción científica tanto en Medicina como en Salud Pública en los contextos internacionales y regionales, mientras que en impacto está entre los últimos países. La tendencia de la producción es al crecimiento, aunque en Salud Pública es más acelerado. El liderazgo es alto, pero la colaboración internacional está por debajo de lo esperado. La publicación en revistas de alto impacto (primer cuartil) y los artículos en el 10% más citado (excelencia) son escasos. Se concluye que el volumen y el impacto de la publicación no están acorde al potencial científico de salud cubana. Se recomienda incrementar la colaboración científica, la publicación de artículos en revistas de alto impacto, la preparación de los recursos humanos y seguir las recomendaciones internacionales sobre las buenas prácticas de edición y publicación científica.

Palabras clave: Cienciometría. Colaboración científica. Evaluación científica. Indicadores bibliométricos. Medicina. Producción científica. Salud colectiva.

Abstract

The aim of this study was to analyze the evolution of the quantity and visibility of Cuban scientific production in Public Health and Medicine to determine if they follow the same patterns of scientific communication and the recommended best practices for publication. Bibliometric indicators of quantity, visibility and cooperation were extracted from the SCImago Institutions Rankings website, which is based on Scopus data, in the field of Medicine and category of Public Health, Environmental and Occupational Health from 2003 to 2011. Cuba has a relatively high position in the rankings of scientific production in both Medicine and Public Health within the international and regional contexts, but its impact is ranked among the last countries. The production trend of both fields has increased, but public

¹ Trabajo presentado en el VII Seminario Internacional sobre Estudios Cuantitativos y Cualitativos de la Ciencia y la Tecnología "Prof. Gilberto Sotolongo Aguilar" en XIII Congreso Internacional de Información - INFO'2014. Habana, Cuba.

² Universidad Virtual de Salud, Centro Nacional de Información de Ciencias Médicas-Infomed, Área de Docencia e Investigaciones. Calle 23, 654 entre D y E, Vedado, 10400, La Habana, Cuba. Correspondencia a nombre de/Correspondence to: G. ZACCA-GONZÁLEZ. E-mail: <grisel.zacca@infomed.sld.cu>.

³ Universidad de Granada, Facultad de Comunicación y Documentación, Departamento de Información y Comunicación, SCImago Research Group. Granada, España.

⁴ Instituto de Políticas y Bienes Públicos, Departamento de Ciencia e Innovación, Consejo Superior de Investigaciones Científicas, SCImago Research Group. Madrid, España.

Recibido el día 29/5/2014 y aceptado para su publicación el 20/8/2014.

health is increasing faster. Leadership is high, but international collaboration is below expectations. Publication in high impact journals (first quartile) and articles in the set 10% most cited documents (excellence) are scarce. Thus, it may be concluded that the volume and impact of publication are not in accordance with the scientific potential of Cuban health. We recommend increasing scientific cooperation, publishing articles in high impact journals, training human resources and following the international recommendations for good editorial and scientific publication practices.

Keywords: Scientometrics. Scientific cooperation. Scientific evaluation. Bibliometric indicators. Medicine. Scientific production. Collective health.

Introducción

En el discurso que ofreciera el ministro de salud cubano como presidente de la 67ª Asamblea Mundial de la Salud, en mayo de 2014, señala:

El Sistema Nacional de Salud cubano se caracteriza por ser único, gratuito, accesible, con cobertura universal, basado en la Atención Primaria de Salud y se distingue por una amplia participación comunitaria e intersectorial y profunda concepción internacionalista. Ejemplos de estos resultados lo constituyen la tasa de mortalidad infantil lograda en el año 2013, de 4.2 por cada mil nacidos vivos, indicador que se ha mantenido por más de cinco años consecutivos por debajo de 5, así como la contribución de Cuba a la salud en más de 120 países del mundo, con la presencia desde 1960 de 135 mil colaboradores y actualmente por encima de 50 mil en 65 países (Morales Ojeda, 2014)⁵.

En este contexto, la investigación es una actividad prioritaria en todas las instituciones de salud, junto a las labores asistenciales y docentes, constituyendo la función fundamental de diferentes centros. Entre ellos destacan los 12 institutos con que cuenta el país, el Polo Científico con sus más de 30 entidades y la Escuela Nacional de Salud Pública, entre otras. La investigación que se desarrolla en el Sistema Nacional de Salud, debe responder a las prioridades y metas de las Áreas Estratégicas de las Proyecciones de la Salud Pública (Domínguez-Alonso & Zacca-Peña, 2011).

Según el informe presentado en La primera Conferencia Latinoamericana sobre Investigación e Innovación para la Salud (Alger *et al.*, 2009), Cuba poseía estructuras formales de gobernanza y gerencia de la investigación para la salud. Aunque no tenían una política nacional específica, contaba con un cuerpo de leyes y

regulaciones que establecían normas y estándares que regulan algunos elementos de la investigación para la salud. Tenía un claro proceso de priorización de las investigaciones en salud, mecanismos de financiamiento, disponía de fondos internacionales para la investigación en salud y tenía una estrategia nacional para la formación de recursos humanos para la investigación en salud.

Los estudios cuantitativos han ganado popularidad como complemento de la evaluación de la actividad científica bajo el presupuesto teórico de que la esencia de la investigación científica es la contribución de nuevos conocimientos al arsenal de conocimientos existentes y que puede ser medida a partir de los artículos publicados en canales de comunicación científica.

Teniendo en cuenta esta premisa se han realizado una serie de estudios, de los cuales éste forma parte, que han caracterizado el dominio de la Salud Pública latinoamericana desde la perspectiva cuantitativa. Se ha identificado a Cuba en el cuarto lugar en volumen de producción científica en Latinoamérica, también como el país con menores tasas de colaboración, entre los países que menos publican en inglés y de los que más artículos tienen en el cuarto cuartil. Los autores cubanos aparecen como los principales de sus publicaciones, por tanto, el liderazgo es alto pero el impacto de la producción científica es bajo, demostrado por el impacto normalizado, bajo el porcentaje de documentos entre el 10% más citado dentro de la categoría (indicador de excelencia) y por bajos valores del indicador de excelencia con liderazgo (Zacca-González *et al.*, 2014; Chinchilla-Rodríguez *et al.*, 2014b).

Otros estudios bibliométricos sobre las ciencias médicas y de la salud han demostrado que la publicación está por debajo de la capacidad investigadora del sistema. Existe una falta de correspondencia importante

⁵ Morales-Ojeda R. Discurso del presidente de la Asamblea Mundial de la Salud, Roberto Morales Ojeda, ministro de Salud Pública de Cuba. In: *Asamblea Mundial de la Salud*, 67., Ginebra, 2014. Disponible en: <files.sld.cu/editorhome/files/2014/05/discurso-del-presidente-de-la-asamblea-mundial-de-la-salud-mayo-2014.pdf>. Acceso en: 11 ago. 2014.

entre los recursos humanos, la infraestructura disponible y el volumen y la visibilidad de publicación científica de Cuba en salud (Dorta Contreras, 2006; Cañedo Andalia, 2013; Cañedo Andalia *et al.*, 2013).

El objetivo de este trabajo es analizar la evolución del volumen y la visibilidad de la producción científica cubana en Salud Pública y en Medicina para determinar si siguen los mismos patrones de comunicación, y en la medida de lo posible, recomendar buenas prácticas tanto a nivel académico como editorial para incrementar la visibilidad de la investigación cubana en Salud Pública.

Métodos

La información bibliométrica se extrajo del portal (SCImago Institutions Rankings, 2014) a partir de datos de Scopus, para el área temática *Medicine* y la categoría *Public Health, Environmental and Occupational Health* que es un subconjunto de la Medicina, en el período 2003-2011.

Los indicadores bibliométricos utilizados fueron:

- Número de documentos: número total de documentos en los que al menos uno de los autores aparece bajo la afiliación de una institución cubana.

- Por ciento de documentos: porcentaje de documentos respecto al total según pertenezca a la Salud Pública o a la Medicina.

- Tasa de crecimiento: diferencia porcentual del número de trabajos en relación al período anterior.

- Colaboración internacional: porcentaje de la producción publicada en colaboración con instituciones de fuera del país.

- Impacto Normalizado: número relativo de citas recibidas por cada país comparado con el promedio mundial de citas por documento de la misma tipología documental, año y categoría. Se calcula utilizando la metodología "*Item oriented field normalized citation score average*" establecida por el *Karolinska Institutet de Suecia* (Rehn *et al.*, 2007) donde la normalización de los valores de citación se hace en nivel del artículo individual. Los valores (en %) muestran las relaciones entre el impacto científico medio de un país y el conjunto promedio mundial con una puntuación de 1. Una puntuación de

0.8 significa que el país es citado un 20% por debajo del promedio mundial y un valor de 1.3 que es citado un 30% superior a la media del mundo.

- *SCImago Journal Rank* (SJR): este indicador se basa en la transferencia de prestigio o influencia desde una revista hacia otra o hacia ella misma a través de referencias. Se calcula con una ventana de citación de 3 años y restringe las autocitas a la revista a un máximo del 33%; se construye con un algoritmo similar al "*Page Rank*" de *Google* (González-Pereira *et al.*, 2010). Este indicador se utiliza para la distribución de las revistas en cuatro cuartiles (Q1, Q2, Q3, Q4) según el SJR, las ubicadas en el cuartil 1 son las que tienen mayor prestigio.

- % output en Q1: porcentaje de publicaciones en revistas incluidas en el primer cuartil (25%) de categoría Salud Pública ordenadas por el SJR (González-Pereira *et al.*, 2010).

- Excelencia científica: porcentaje de producción científica de un país que se ha incluido en el grupo del 10% de trabajos más citados en el área o categoría temática. Es una medida de la cantidad de producción científica de alta calidad (Bornmann *et al.*, 2012).

- Liderazgo Científico: Porcentaje de producción de un país en el que el autor principal ("corresponding author") pertenece a las instituciones nacionales de cada país. Se le denomina documentos liderados (Moya-Aneón *et al.*, 2013).

- Excelencia con liderazgo: Por ciento de documentos liderados de un país que se encuentra entre el 10% más citado (Jeremić *et al.*, 2013).

Resultados

Cuba en el contexto internacional

Scopus contiene, en el período 2003-2011, una producción científica de 4.488.365 documentos en el área de la Medicina; a este total Cuba aportó el 0,14%. En Salud Pública en el mundo se publicaron 212 310 documentos, lo que representó el 4,73% de la Medicina. Cuba contribuyó con el 0,33% (696 documentos). Para ubicar a Cuba en el contexto internacional utilizamos el ranking mundial de países de acuerdo al volumen de producción

y la visibilidad a partir del indicador Impacto Normalizado (IN) (Tabla 1). Cuba se ubica en el lugar 51 en el ranking por Número de documentos (Ndoc) en Medicina. Esta lista está encabezada por Estados Unidos (26,57%), el Reino Unido (7,76%) y Alemania (6,41%). En el ranking internacional de Salud Pública, Cuba sube hasta el lugar 39. Entre los 20 países mayores productores, el único país latinoamericano que aparece es Brasil en el lugar 14 en Medicina con una contribución de 1,91% y el quinto en Salud Pública con un aporte mayor de 4,43% (Tabla 1).

El ranking de impacto muestra una situación desventajosa para Cuba (Tabla 1). De los 71 países con al menos 2000 documentos en Medicina, Cuba ocupa el último lugar en Medicina y el penúltimo en Salud Pública. No hay ningún país latinoamericano entre los 20 más visibles de acuerdo al impacto normalizado.

Cuba en el contexto regional

Al volumen de producción científica latinoamericana en Medicina de 145.646, la Salud Pública aporta el 9,78% (14.249). Cuba publicó 6.233 documentos en Medicina y 696 en Salud Pública, lo que contribuye a la región con 4,28% y 4,88% respectivamente.

En Medicina, Cuba se posicionó en el sexto lugar en el ranking de países según el volumen de producción del período 2003 -2011; en el análisis de la posición en cada año, inició el período en el quinto lugar y luego fue desplazada por Colombia al sexto lugar en 2008. Los primeros países del ranking fueron Brasil, México y Argentina. En Salud Pública, Cuba se ubicó en el quinto puesto; en el análisis anual no siempre se comportó de esta manera. Ocupó los lugares cuarto (2006, 2007 y 2011), quinto (2003, 2008 y 2010) y sexto (2009) intercambiando posiciones con Argentina, Colombia y Chile (Tabla 2).

En la tabla 2 también se analiza el impacto de las publicaciones de los 20 países latinoamericanos con mayor producción. Tanto en la Medicina como en la Salud Pública, Cuba se ubicó al final del ranking. En este ranking llama la atención la posición de países con poca producción y con un alto impacto, que en la mayoría de los casos es debido a las colaboraciones internacionales que establecen los países con menor producción (Chinchilla-Rodríguez *et al.*, 2012).

Evolución de la producción científica

En Cuba, la contribución que hace la Salud Pública al área temática de la Medicina fue de 11,12%. Ambos dominios tienden al crecimiento de su producción científica, sin embargo, el incremento de la Salud Pública (223,8%) fue más rápido y superó en 65 puntos a la Medicina (158,6 %). Este comportamiento también se reflejó en el ritmo de crecimiento por año, dado que la tasa de crecimiento anual promedio de la Medicina fue de 15,71% y la de la Salud Pública de 21,72%. Al inicio del período la contribución de la Salud Pública a la Medicina fue de 12,36% y ya en 2011 asciende a 15,11%.

En la comparación de la evolución de ambos agregados se observaron ciertas diferencias. Hubo coincidencia en el mayor incremento entre 2005 y 2006 y disminución en 2008. La Salud Pública disminuyó entre 2004 y 2005 y la Medicina lo hizo en 2010 (Figura 1).

La Figura 2 muestra la comparación de los indicadores bibliométricos de la Medicina y la Salud Pública: colaboración internacional, impacto normalizado, por ciento de documentos en el primer cuartil, la excelencia (producción científica de alto impacto), el liderazgo y la excelencia con liderazgo. En términos generales, se apreció un comportamiento variable de todos los indicadores con períodos de alzas y bajas.

El porcentaje de producción científica en colaboración internacional de la Salud Pública fue de 19,3%, ligeramente superior al de la Medicina (18,2%), sin embargo, a pesar de que ambos agregados tienen porcentajes similares, en la Salud Pública se apreció una ligera tendencia a la disminución, al contrario de la Medicina que varió menos. Estas bajas tasas de colaboración están acompañadas de un alto porcentaje de liderazgo en ambos dominios. Es por ello que acompañado de la disminución de la colaboración, la tendencia de la Salud Pública es el incremento de los documentos liderados, mientras que la Medicina tiende a mantenerse estable.

El por ciento de producción en el primer cuartil (25% según SJR) para la Salud Pública fue inferior al de la Medicina, en ambos dominios tiende a decrecer, característica que es un poco más marcada en la Salud Pública; ambos agregados tuvieron un alza en 2008 y la

Tabla 1. Ranking mundial de países organizados por volumen de producción y visibilidad de las publicaciones en Medicina y Salud Pública, 2003-2011.

Medicina					
Rank	País	Ndoc	Rank	País*	IN
1	Estados Unidos	1.192.666	1	Islandia	2,55
2	Reino Unido	348.617	2	Kenia	2,10
3	Alemania	287.722	3	Dinamarca	2,02
4	Japón	245.704	4	Uganda	2,02
5	China	232.312	5	Holanda	1,94
6	Francia	187.346	6	Finlandia	1,93
7	Italia	181.750	7	Bélgica	1,89
8	Canadá	161.465	8	Suecia	1,87
9	España	139.865	9	Suiza	1,86
10	Australia	124.252	10	Canadá	1,84
11	Holanda	112.678	11	Tanzania	1,81
12	India	91.126	12	Noruega	1,78
13	Turquía	87.907	13	Perú	1,71
14	Brasil	85.668	14	Estados Unidos	1,66
15	Suiza	71.253	15	Australia	1,65
16	Corea del Sur	64.972	16	Reino Unido	1,64
17	Suecia	63.200	17	Nueva Zelanda	1,64
18	Polonia	57.168	18	Austria	1,59
19	Bélgica	53.439	19	Singapur	1,57
20	Taiwán	47.709	20	Irlanda	1,56
51	Cuba	6.233	71	Cuba	0,27
Salud Pública					
Rank	País	Ndoc	Rank	País	IN
1	Estados Unidos	73.876	1	Suiza	1,96
2	Reino Unido	20.031	2	Dinamarca	1,90
3	Canadá	11.571	3	Noruega	1,70
4	Alemania	9.525	4	Holanda	1,69
5	Brasil	9.413	5	Kenia	1,56
6	Australia	9.386	6	Grecia	1,55
7	Italia	6.372	7	Finlandia	1,54
8	Francia	6.013	8	Irlanda	1,49
9	España	5.748	9	Tanzania	1,48
10	Holanda	5.486	10	Bélgica	1,47
11	China	4.361	11	Suecia	1,46
12	Suecia	4.240	12	Canadá	1,42
13	Japón	3.520	13	Reino Unido	1,40
14	India	3.495	14	Estados Unidos	1,37
15	Suiza	3.383	15	Singapur	1,33
16	Dinamarca	2.242	16	Uganda	1,33
17	Noruega	2.230	17	Australia	1,32
18	Bélgica	2.226	18	Nueva Zelanda	1,30
19	Suráfrica	2.156	19	Hungría	1,30
20	México	2.049	20	Portugal	1,27
39	Cuba	696	70	Cuba	0,24

Fuente: SCImago Institutions Rankings (2014). Elaboración propia.

Nota: *se incluyeron 71 países con 2 000 documentos o más en Medicina, Ndoc: Número de documentos; IN: Impacto normalizado.

Tabla 2. Ranking latinoamericano de países organizados por volumen de producción y visibilidad de la publicación en Medicina y Salud Pública, 2003-2011.

Medicina					
Rank	País	Ndoc	Rank	País	IN
1	Brasil	85.668	1	Costa Rica	1,86
2	México	22.751	2	Panamá	1,73
3	Argentina	15.349	3	Perú	1,71
4	Chile	10.196	4	Barbados	1,60
5	Colombia	6.780	5	Ecuador	1,34
6	Cuba	6.233	6	Bolivia	1,34
7	Venezuela	3.861	7	Puerto Rico	1,33
8	Perú	2.452	8	Argentina	1,28
9	Puerto Rico	2.107	9	Uruguay	1,19
10	Jamaica	1.859	10	Guatemala	1,18
11	Uruguay	1.383	11	Trinidad y Tobago	0,95
12	Costa Rica	911	12	Chile	0,93
13	Ecuador	811	13	Jamaica	0,92
14	Trinidad y Tobago	710	14	Colombia	0,86
15	Guatemala	480	15	Brasil	0,85
16	Granada	434	16	México	0,81
17	Panamá	343	17	Venezuela	0,74
18	Barbados	321	18	Martinica	0,68
19	Bolivia	315	19	Granada	0,54
20	Martinica	315	20	Cuba	0,27

Salud Pública					
Rank	País	Ndoc	Rank	País	IN
1	Brasil	9.413	1	Bolivia	1,62
2	México	2.049	2	Uruguay	1,36
3	Colombia	874	3	Guatemala	1,36
4	Chile	754	4	Perú	1,23
5	Cuba	696	5	Panamá	1,17
6	Argentina	635	6	Ecuador	1,08
7	Perú	315	7	Argentina	1,05
8	Venezuela	205	8	Venezuela	1,03
9	Puerto Rico	203	9	Granada	1,01
10	Jamaica	182	10	Puerto Rico	0,90
11	Costa Rica	128	11	Costa Rica	0,88
12	Ecuador	110	12	Barbados	0,86
13	Uruguay	106	13	México	0,81
14	Trinidad y Tobago	90	14	Jamaica	0,77
15	Guatemala	84	15	Brasil	0,72
16	Bolivia	63	16	Trinidad y Tobago	0,68
17	Barbados	44	17	Chile	0,60
18	Panamá	29	18	Colombia	0,59
19	Granada	16	19	Martinica	0,30
20	Martinica	9	20	Cuba	0,24

Fuente: SCImago Institutions Rankings (2014). Elaboración propia.

Nota: del total de países latinoamericanos se incluyeron en este ranking los 20 primeros según número de documentos.

Ndoc: Número de documentos; IN: Impacto normalizado.

Salud Pública además en 2006. El porcentaje de documentos, en el 10% de mayor impacto (excelencia) fue escaso en el área temática y la categoría y, a pesar de la irregularidad en el comportamiento de este indicador

(sobre todo en Salud Pública) se puede apreciar una ligera tendencia al incremento del indicador en la Salud Pública y a la disminución en la Medicina; la Salud Pública no tuvo ningún documento en este conjunto en los dos

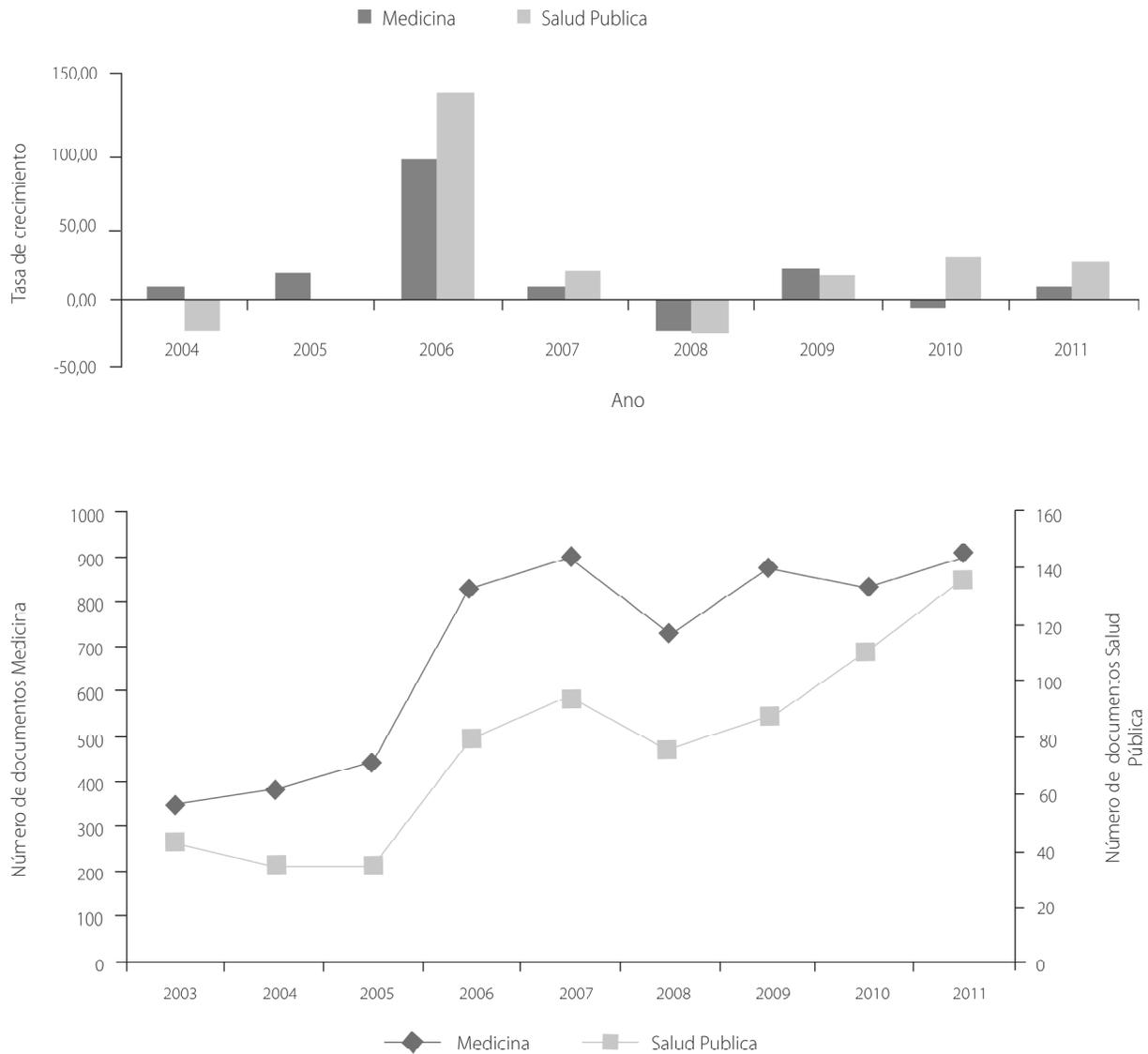


Figura 1. Número de documentos y tasa de crecimiento de la producción científica cubana en Medicina y Salud Pública, 2003-2011.

Fuente: SCImago Institutions Rankings (2014). Elaboración propia.

primeros años del período y se mantuvo por debajo de la Medicina excepto en 2005, 2007 y 2008. Como es tan baja la presencia en este conjunto de alta calidad, el % de documentos que tienen autores cubanos como autor de la correspondencia en la Salud Pública es casi inexistente, manteniéndose en cero dos tercio de los años del período y alcanzando un máximo de 2,86% en 2005; en la Medicina es bajo el % de documentos de excelencia con liderazgo y tiende a la disminución.

El indicador del impacto normalizado, que compara la visibilidad del país con las medias mundiales,

posiciona a Cuba lejos de los estándares internacionales. Las citas obtenidas por ambos dominios estuvieron la mayor parte del período cerca de un 70% por debajo de la media mundial, llegando a estar por debajo del 80%; el valor más alto lo obtuvo la Medicina al inicio del período (50% por debajo del mundo). Los valores del indicador evidenciaron una tendencia al descenso del impacto de la producción contenida en ambos dominio. La Salud Pública solo superó al área temática en 2005, y a partir del 2007 tuvieron un comportamiento similar dado por un incremento en 2008 y 2010.

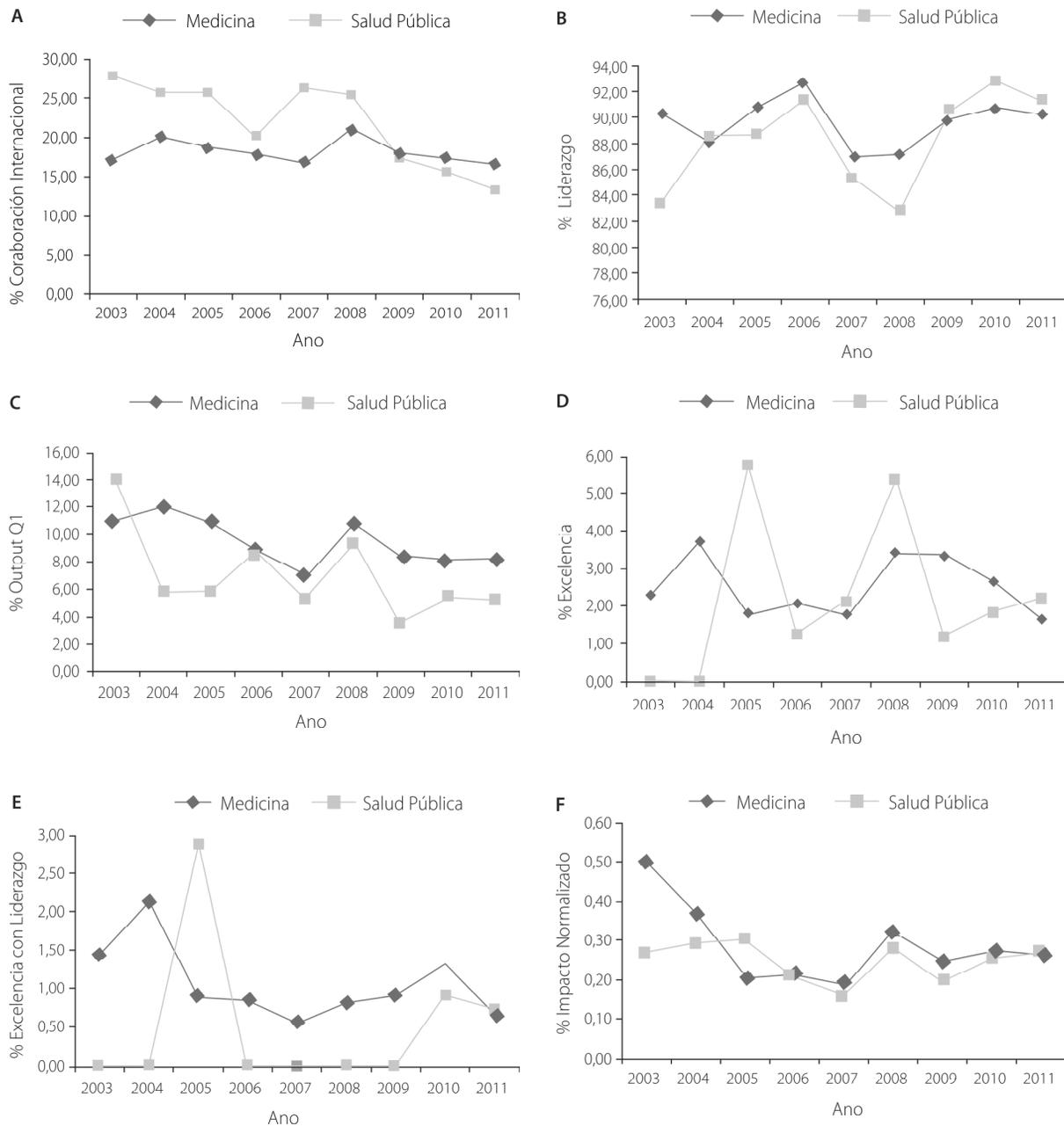


Figura 2. Indicadores bibliométricos de la producción científica cubana en Medicina y Salud Pública, 2003-2011.

Fuente: SCImago Institutions Rankings (2014). Elaboración propia.

Nota: A: Colaboración internacional; B: Liderazgo; C: % de producción en Q1; D: Excelencia; E: Excelencia con Liderazgo; F: Impacto normalizado.

Publicación en revistas cubanas y extranjeras en el área de la Medicina y la categoría Salud Pública

En relación a la distribución de las publicaciones de autores cubanos por revistas y el impacto que éstas alcanzan (Tabla 3) según el cuartil (Q) donde se ubican,

el mayor volumen de producción se ubica en Q4 en ambos agregados, aunque en proporciones mayores en Salud Pública (55%) que en Medicina (42%) y con una clara tendencia al crecimiento; ambos dominios experimentaron el mayor crecimiento a partir de 2006, fundamentalmente en la Salud Pública que ubicó en este

Tabla 3. Distribución e impacto de la producción científica en Salud Pública y Medicina según los cuartiles (Q) de las revistas donde publican autores cubanos, ordenados según el SCImago Journal Rank, 2003-2011.

Año	Salud Pública								Medicina							
	Q4		Q3		Q2		Q1		Q4		Q3		Q2		Q1	
	%	IN	%	IN	%	IN	%	IN	%	IN	%	IN	%	IN	%	IN
2003			7,14	0,09	73,81	0,28	19,05	0,79	10,63	0,27	41,09	0,08	33,05	0,21	15,23	2,81
2004			57,14	0,06	37,14	0,44	5,71	1,01	9,26	0,15	44,71	0,08	29,10	0,21	17,99	1,41
2005			8,57	0,02	85,71	0,82	5,71	1,00	8,39	0,15	28,34	0,04	51,02	0,09	12,70	1,28
2006	57,50	0,01	20,00	0	13,75	0,73	8,75	1,18	17,92	0,02	56,30	0,01	16,34	0,26	10,77	1,65
2007	53,68	0	31,58	0,16	9,47	0,52	5,26	1,88	37,22	0,01	43,78	0,03	13,56	0,28	7,89	1,75
2008	74,67	0,01	5,33	0,13	6,67	0,37	13,33	1,86	51,71	0,02	29,96	0,08	11,90	0,34	12,45	2,01
2009	83,53	0,04	2,35	0,34	10,59	1,07	3,53	1,79	68,80	0,04	10,86	0,15	12,46	0,43	9,60	1,79
2010	84,55	0,04	0,91	0,36	9,09	0,24	5,45	3,37	61,34	0,02	16,09	0,14	14,05	0,34	8,88	2,17
2011	81,62	0,04	2,21	0,52	11,03	0,44	5,15	3,30	63,56	0,02	17,78	0,12	9,67	0,57	9,22	1,94

Fuente: SCImago Institutions Rankings (2014). Elaboración propia.

Nota: IN: Impacto normalizado.

cuartil más de tres cuartas partes de su total de documento. Al inicio del período (2003-2005), y previo al incremento, los artículos se concentraron fundamentalmente en revistas del segundo y tercer cuartil en ambos dominios. En Q1 la publicación es escasa, aunque mayor en Medicina que en la categoría. La tendencia en ambos dominios es al incremento de la publicación en Q4 y disminución en el resto de los cuartiles.

El mayor impacto lo obtienen los trabajos publicados en Q1; aunque los trabajos publicados en el área de la Medicina alcanzan mayor impacto que la categoría, la Salud Pública mostró una tendencia mayor al incremento del impacto normalizado alcanzado en este cuartil. Los artículos en Q2 no alcanzan la media internacional aunque se aprecia cierta tendencia a la mejora en ambos agregados. En Q3 el impacto está por debajo del 60% y ambos tienden al crecimiento. El impacto de Q4 es muy bajo en ambos dominios y la tendencia es a permanecer en los niveles más bajos.

Del total de producción científica, el 63,53% está publicado en 14 revistas nacionales de las 19 revistas médicas cubanas que aparecen en Scopus. Las que contienen más de 300 documentos y aportan más del 5% al total de producción son: Revista Cubana de Medicina General Integral (7,67%), Acimed (7,36%), Revista Cubana de Medicina Tropical (5,50%), Revista Cubana de Medicina (5,58%) y la Revista Cubana de Cirugía (5,49%).

Las revistas indizadas en la categoría de Salud Pública son la Revista Cubana de Salud Pública y la Revista Cubana de Higiene y Epidemiología, las que contribuyeron al área temática con un 3,98% y 3,67% respectivamente. En su conjunto aportaron el 7,65% al total de producción en Medicina y el 68,53% a la Salud Pública.

Discusión

Estudios anteriores han demostrado que la utilización de Scopus como fuente para análisis cuantitativos, evidencia la orientación biomédica de la investigación cubana (Arencibia Jorge & Moya Anegón, 2010). Cuba es un país relativamente pequeño, con alrededor de 11 millones de habitantes, que tiene una alta proporción de publicaciones por millón de habitantes (Chinchilla-Rodríguez *et al.*, 2014b). Es por ello que se considera relevante que ocupe el lugar 51 en el área temática Medicina y el 39 en la categoría Salud Pública en el ranking internacional de países por volumen de producción. Dentro de la Medicina, la Salud Pública constituye una fortaleza, lo que en cierta manera refleja el quehacer salubrista de la isla, aunque en comparación con los logros obtenidos, el volumen de producción científica, que suponen ser un reflejo de los resultados de la investigación, está por debajo de las capacidades existentes.

Un ejemplo de esto es la escasa publicación en Atención Primaria en Salud, base de la práctica médica

cubana. Al respecto (Benet, 2013) planteó que es una situación preocupante si se considera que la residencia de la medicina familiar, especialidad que realizan la mayoría de los médicos recién graduados, termina con una tesis y que las maestrías y doctorados en Salud Pública y otros campos relacionados, también contribuyen al potencial de investigación. El análisis de la situación de salud de la comunidad que se realiza anualmente no está suficientemente reflejado en la literatura.

En el análisis de la evolución de la producción científica en Medicina y su subconjunto Salud Pública, se observa una tendencia al incremento del volumen total, que es más acelerado en la Salud Pública. Esta producción, altamente liderada, experimentó, el mayor crecimiento entre 2005-2006 a expensas del incremento de las revistas en el cuarto cuartil, coincidiendo con la incorporación de las revistas cubanas indizadas en *SciELO* a *Scopus*, las que inmediatamente pasaron a formar parte de Q4. Este aumento de producción ha repercutido en una disminución el porcentaje de documentos publicados en las mejores revistas (Q1) en la mayor parte del período, excepto una disminución en ambos agregados en 2008 en Q4 asociado al incremento de publicaciones en revistas del primer cuartil. La baja proporción de artículos de excelencia y la casi inexistente cantidad de documentos de excelencia liderados por autores cubanos, sitúan a Cuba con una citación muy alejada del promedio mundial, y por extensión con baja visibilidad e impacto a nivel internacional. Estos resultados coinciden con los descritos por (Chinchilla-Rodríguez *et al.*, 2014a) en el estudio de la producción cubana en todas las áreas de conocimientos contenidas en *Scopus*.

Se precisó una tendencia al incremento de documentos en Q4 y decrecimiento en el resto de los cuartiles en ambos dominios. No obstante, en Medicina parece haber una mayor estabilidad que en Salud Pública en cuanto a las contribuciones que se hacen en revistas del primer cuartil. Este es un elemento que alerta sobre la necesidad de incrementar la publicación de revistas de alto impacto, particularmente en Salud Pública.

Tanto en Medicina como es Salud Pública, hay un desbalance hacia la publicación mayor en revistas

nacionales que extranjeras, lo que confirma el patrón hallado para las ciencias médicas y de la salud por (Cañedo Andalia *et al.*, 2013). Por tanto, como consecuencia más inmediata del aumento del volumen de la literatura científica a partir de la introducción de las revistas médicas cubanas a *Scopus* se ha producido un descenso de la citación. No obstante, este fenómeno debe ser observado durante los próximos años, para dar suficiente tiempo a que estos artículos sean citados y poder realizar un análisis objetivo de la calidad de las investigaciones (Arencibia-Jorge & Moya-Anegón, 2010).

Por tanto, la capacidad que tiene Cuba para publicar en el área de la Medicina no se manifiesta en el nivel de impacto alcanzado por la producción científica. Otra evidencia de esto fue la ubicación en el último lugar en los rankings de países de Medicina y Salud Pública, tanto en el contexto regional como internacional, posiciones que contrastan grandemente con la ubicación en cuanto a volumen de producción.

Las 14 revistas cubanas que forman parte del área de la Medicina pertenecen al Sistema Nacional de Salud y son publicadas por la Editorial de Ciencias Médicas, Ecimed. Circulan en acceso abierto en el sitio de las revistas médicas cubanas en la Biblioteca Virtual de Salud de Cuba y en *SciELO*. Su incorporación a *Scopus* ha permitido dar visibilidad internacional a más de seis mil artículos en los 9 años estudiados.

El hecho de que estas revistas circulen por estos índices internacionales presupone el cumplimiento de los estándares de calidad de *SciELO* y *Scopus*. Sin embargo, estar en acceso abierto no necesariamente conlleva a una mayor citación. Se ha demostrado que estas revistas de la vía dorada del *Open Access*, tienen poca visibilidad y se sitúan en el cuarto cuartil (Miguel *et al.*, 2011).

La colaboración internacional aparece por debajo de lo esperado en ambos dominios teniendo en cuenta el número de profesores y especialistas cubanos que prestan servicios de colaboración médica en otros países del mundo, principalmente en Latinoamérica. La tendencia de la colaboración científica en Salud Pública es a disminuir mientras que la de la Medicina se mantiene constante. Una mayor apertura internacional tanto con países de la región como de fuera de la región podría ser

una buena recomendación (Lancho-Barrantes *et al.*, 2012; 2013). Esta estrategia de mayor liderazgo científico, unido a la presencia de revistas en bases de datos internacionales, indica que las condiciones están creadas para establecer redes de colaboración y alcanzar mayor repercusión y visibilidad de las investigaciones científicas cubanas más genuinas y de esta manera pudiera contribuirse al desbloqueo académico. En la realidad actual de las ciencias médicas cubanas, esta estrategia de apertura es crucial por la necesidad de respaldar la labor médica de los colaboradores cubanos en 65 países del mundo, donde la situación de salud difiere de la cubana.

Cuba es un país en vías de desarrollo donde la salud de la población es una prioridad. Ha alcanzado altos niveles de salud, comparables a los de países desarrollados y es el país de Latinoamérica con mayor equidad en términos salubristas (Chinchilla-Rodríguez *et al.*, 2014b). La inversión en salud es relativamente alta comparada con el Producto Interno Bruto (PIB) y cuenta con un alto potencial de recursos humanos altamente calificados que ha sido denominado "ejército de batas blancas". Los perfiles de los currículos de especialidad y maestría tienen salida en la investigación y se les exige a los docentes e investigadores categorizados un número determinado de publicaciones anuales según categoría. Sin embargo, como se ha demostrado en esta investigación, estas potencialidades en recursos humanos e infraestructura del Sistema Nacional de Salud que pudiera ser asiento de numerosas investigaciones en Medicina, y especialmente en Salud Pública, no se corresponden con la cantidad y calidad de publicaciones visibles internacionalmente.

Dorta Contreras ha destacado que en Cuba la producción científica en las ciencias médicas y de la salud está por debajo del potencial humano existente, por factores tales como la falta de cultura editorial de los profesionales, insuficiente formación en materia de redacción de documentos científicos, la existencia de prioridades como la docencia médica y la asistencia que conspira contra el tiempo disponible para la preparación de las publicaciones; dificultades para seleccionar correctamente una revista y el pobre manejo del idioma inglés (Dorta Contreras, 2006). También atribuye la falta

de citas entre cubanos a que no existe integración en la ciencia nacional, la falta de coherencia temática de los autores, que no se considera importante la contribución que realizan los colegas nacionales o que se desconozca el tema de estudio de éstos (Dorta Contreras, 2008).

Es, por tanto, necesario sistematizar y publicar las investigaciones que dan origen a la toma de decisiones y a documentar la puesta en marcha de iniciativas y programas de salud exitosos que pudieran ser utilizados por la comunidad científica. Por ejemplo, incrementar el número de artículos que recojan los resultados del enfrentamiento a la epidemia del dengue que afecta a varios países de la región o las experiencias de programas como el materno-infantil.

No debe desestimarse la influencia que puede tener en la visibilidad, la publicación en revistas de prestigio internacional como canales de difusión académica de las investigaciones con mejores diseños y resultados generalizables. Es necesario establecer una estrategia para la publicación en revistas de mayor prestigio, buscando siempre un balance entre los artículos de alta calidad que se publican en revistas cubanas y las que se publican en revistas extranjeras, así como priorizar la publicación en revistas multilingües.

Por otra parte, se precisa incrementar las acciones capacitantes en materia de metodología de la investigación y publicación científica. Los programas académicos que contienen estos temas, así como los que forman parte del programa de alfabetización informacional, deben incluir entre sus objetivos educativos lo relacionado con la evaluación de las revistas para que los autores tomen decisiones informadas a la hora de seleccionar la revista donde van a publicar los resultados de sus investigaciones. Asimismo, debe ser prioridad seguir trabajando en la mejora de la calidad de las revistas médicas cubanas, fundamentalmente en el proceso de arbitraje por pares, la composición de sus comités editoriales y en la salida en tiempo de las revistas.

Para el incremento del volumen y visibilidad de las revistas indexadas en Scopus se sugiere realizar estudios de cobertura de las bases de datos y revisión de que estén indizados todos los números de las revistas; elevar la calidad y visibilidad de las revistas cubanas mediante el perfeccionamiento de los procesos

editoriales, fundamentalmente, el arbitraje por pares; e invitar a expertos internacionales a publicar en las revistas nacionales.

Conclusión

Cuba hace un aporte importante a la producción científica en Medicina, especialmente al arsenal de resultados de investigación en Salud Pública. Esta contribución altamente liderada, evidencia una tendencia al incremento en los índices internacionales. Sin embargo, esta vasta producción científica no está impactando en la comunidad internacional como se esperaría de un país con grandes logros en materia de salud. La situación de bajo impacto de los artículos de Medicina es aún más crítica en el subconjunto que forma parte de la Salud Pública. Entre los factores comunes que contribuyen a que ambos dominios tengan bajo impacto

pueden precisarse el bajo porcentaje de colaboración internacional, la alta publicación en revistas de poco prestigio internacional y la escasa presencia de artículos en revistas de alto impacto o trabajos poco generalizables para la comunidad internacional.

Las claves para mejorar el impacto de la producción cubana en Medicina y Salud Pública son el incremento la colaboración científica, la publicación de los artículos en revistas de alto impacto, la preparación de los recursos humanos y seguir las recomendaciones internacionales sobre las buenas prácticas de edición y publicación científica.

Este estudio alerta sobre la necesidad de profundizar en el patrón de comunicación científica de la Salud Pública cubana en aras de determinar qué elementos dan al traste con la calidad de la publicación científica en este dominio de conocimiento tan importante para la nación cubana.

Referencias

Alger, J. *et al.* Sistemas nacionales de investigación para la salud en América Latina: una revisión de 14 países. *Revista Panamericana de Salud Pública*, v.26, n.5, p. 447-57, 2009.

Arencibia-Jorge, R.; Moya-Anegón, F. Challenges in the study of Cuban scientific output. *Scientometrics*, v.83, n.3, p.723-737, 2010.

Benet, M. Cuban publishing on primary health care: An inexcusable absence. *MEDICC Review*, v.15, n.2, p.52, 2013.

Bornmann, L.; Moya-Anegón, F.; Leydesdorff, L. The new excellence indicator in the world report of the SCImago Institutions Rankings 2011. *Journal of Informetrics*, v.6, n.2, p.333-335, 2012.

Cañedo Andalia, R. Se reinicia en PubMed-Medline el procesamiento bibliográfico de la Revista Cubana Medicina Tropical. *Correo Científico Médico*, v.17, n.2, p. 242-243, 2013.

Cañedo Andalia, R.; Rodríguez Labrada, R.; Velázquez Pérez, L. Distribución de la producción científica cubana en salud registrada en Scopus y PubMed en 2011, según instituciones. *Revista Cubana de Información en Ciencias de la Salud*, v.24, n.1, 2013. Disponible en: <<http://www.acimed.sld.cu/index.php/acimed/article/view/388/288>>. Acceso en: 20 dic. 2013.

Chinchilla-Rodríguez, Z. *et al.* International collaboration in medical research in Latin America and the Caribbean (2003-2007). *Journal of the American Society for Information Science and Technology*, v.63, n.11, p.2223-2238, 2012.

Chinchilla-Rodríguez, Z. *et al.* *Colaboración y performance científico en el dominio científico de Cuba en Scopus, 2003-2011.*

In: Congreso Internacional de Información Info'2014, 23., La Habana. *Proceedings...* La Habana: Instituto de Información Científica y Tecnológica, 2014a. 1 CD-ROM.

Chinchilla-Rodríguez, Z. *et al.* Latinoamerican scientific output in Public Health: combined analysis of bibliometrics, socioeconomic and health indicators. *Scientometrics*, 2014b. Available from: <<http://link.springer.com/article/10.1007/s11192-014-1349-9>>. Cited: Aug. 11, 2014b. doi: 10.1007/s11192-014-1349-9

Domínguez-Alonso, E.; Zacca-Peña, E. Sistema de salud de Cuba. *Salud Pública México*, v.53, n.2, p.168-176, 2011.

Dorta Contreras, A. En defensa de nuestra producción científica. *Acimed*, v.14, n.3, 2006. Disponible en: <http://bvs.sld.cu/revistas/aci/vol14_3_06/aci15306.htm>. Acceso en: 15 dic. 2013.

Dorta Contreras, A. Evidenciar la ciencia cubana. *Acimed*, v.17, n.2, 2008. Disponible en: <http://bvs.sld.cu/revistas/aci/vol17_2_08/aci01208.htm>. Acceso en: 15 dic. 2013.

González-Pereira, B.; Guerrero-Bote, V.; Moya-Anegón, F. A new approach to the metric of journal's scientific prestige: The SJR indicator. *Journal of Informetrics*, v.4, n.3, p.379-391, 2010.

Jeremić, V. *et al.* Excellence with leadership: The crown indicator of Scimago Institutions Rankings Iber Report. *El Profesional de la información*, v.22, n.5, p.474-480, 2013.

Lancho-Barrantes, B.; Guerrero-Bote, V.; Moya-Anegón, F. Citation increments between collaborating countries. *Scientometrics*, v.94, n.3, p.817-831, 2013.

Lancho-Barrantes, B. *et al.* Citation flows in the zones of influence of scientific collaborations. *Journal of the American Society for Information Science and Technology*, v.63, n.3, p.481-489, 2012.

Miguel, S.; Chinchilla-Rodríguez, Z.; Moya-Anegón, F. Open Access and Scopus: A new approach to scientific visibility from the standpoint of access. *Journal of the American Society for Information Science and Technology*, v.62, n.6, p.1130-1145, 2011.

Moya-Anegón, F. *et al.* The research guarantors of scientific papers and the output counting: A promising new approach. *Scientometrics*, v.97, n.2, p.421-434, 2013.

Rehn, C.; Kronman, U.; Wadskog, D. *Bibliometric indicators: Definitions and usage at Karolinska Institutet Version 1.0.*

Estocolomo: Karolinska Institutet University Library, p.1-33, 2007. Disponible en: <http://kib.ki.se/sites/kib.ki.se/files/Bibliometric_indicators_definitions_1.0.pdf>. Acceso en: 11 ago. 2014.

SCImago Institutions Rankings. *Base de datos*. 2014. Disponible en: <<http://www.scimagoir.com>>. Acceso en: 11 mar. 2014.

Zacca-González G. *et al.* Bibliometric analysis of regional Latin America's scientific output in public health through SCImago journal country rank. *BMC Public Health*, v.14, n.632, 2014. Available from: <<http://www.biomedcentral.com/1471-2458/14/632>>. Cited: Aug. 11, 2014. doi 10.1186/1471-2458-14.632

A organização do conhecimento na dinâmica da pesquisa em artigos da literatura científica da Brapci¹

Knowledge organization in the dynamics of research in the scientific literature of articles at Brapci

Juliana Lazzarotto FREITAS²

Bruna Silva do NASCIMENTO³

Leilah Santiago BUFREM²

Resumo

Este artigo analisa as dimensões de pesquisa expressas concretamente na literatura da área de Organização do Conhecimento. Considera que a produção científica desenvolvida em determinados contextos sociais e momentos históricos reflete as mudanças e contradições desse contexto, tanto em sua organização interna quanto em suas aplicações. Questiona como se configuram as relações entre as opções metodológicas adotadas nessa literatura e o processo dinâmico de pesquisa sob a ótica multidimensional. Ilustra esse processo de análise a partir de um *corpus* de 105 artigos recuperados da Base Referencial de Artigos de Periódicos em Ciência da Informação, entre 2003 e 2012. Utiliza os *softwares Endnote* e *Excel* para organização e análise dos dados coletados. Categoriza os artigos de pesquisa sob o ponto de vista de seu objeto e delineamento metodológico, bem como de suas relações teóricas por meio de estudo de citação. Evidencia os autores mais expressivos e atuantes na área de Organização do Conhecimento no Brasil. Identifica os periódicos mais devotados ao tema e conclui que a produção científica brasileira sobre o tema distribui-se de maneira irregular ao longo do período analisado, tendo seu ápice no ano de 2011, durante o qual são publicados 18,27% dos artigos. Finaliza estabelecendo relações entre as correntes teóricas da Organização do Conhecimento e os enfoques e temáticas predominantes no *corpus*. Os enfoques de pesquisa de maior incidência foram o teórico e o linguístico, e o tipo de análise foi documental.

Palavras-chave: Literatura periódica científica. Organização do conhecimento. Pesquisa científica.

Abstract

The aim of this paper is to analyze the research dimensions expressed in the literature in the field of knowledge organization. We consider that scientific investigation developed in certain social contexts and historical conjunctures reflects the changes and contradictions present in these contexts, both in its organization and applications. This process is illustrated from a corpus of 105 articles retrieved from

¹ Trabalho apresentado no VII Seminário Internacional sobre Estudos Quantitativos e Qualitativos da Ciência e Tecnologia "Prof. Gilberto Sotolongo Aguilar" no XIII Congresso Internacional de Informação - INFO'2014. Habana, Cuba.

² Universidade Estadual Paulista Júlio de Mesquita Filho, Faculdade de Filosofia e Ciências, Programa de Pós-Graduação em Ciência da Informação. Av. Hygino Muzzi Filho, 737, Mirante, 17525-000, Marília, São Paulo, SP, Brasil. Correspondência para/Correspondence to: L.S. BUFREM. E-mail: <santiagobuffrem@gmail.com>.

³ Universidade Federal do Estado do Rio de Janeiro, Escola de Biblioteconomia, Departamento de Estudos e Processos Biblioteconômicos. Rio de Janeiro, RJ, Brasil.

Recebido em 2/6/2014 e aceito para publicação em 20/8/2014.

the referential database of journal articles on Information Science between 2003 and 2012. Endnote and Excel software were used for organizing and analyzing the data collected. The articles are categorized according to their object and methodological approach, as well as their theoretical foundations through the study of citations. Throughout this investigation, we point out the most expressive and active authors in the field of knowledge organization in Brazil. We also identify the journals that are more devoted to the theme and we conclude that Brazilian scientific production on the theme is irregularly distributed over the period of time analyzed, peaking in 2011 when 18,27% of the total number of articles were published. We conclude by establishing the relationship between theoretical currents of knowledge organization and the prevalent approaches and themes found in the corpus. The focus of the most prevalent research approaches were the theoretical and linguistic approaches and the type of analysis was documentary.

Keywords: Scientific journal literature. Knowledge organization. Scientific research.

Introdução

A análise de desdobramentos metodológicos das pesquisas em Organização do Conhecimento (OC) no Brasil possibilita a caracterização desse domínio como alternativa para o repensar crítico e autocrítico sobre os meios de construção da pesquisa na área. A percepção da atividade de pesquisa como processo dinâmico, social e histórico - e não somente como recurso para o aprimoramento do fazer científico -, é evidenciada nos argumentos de Lloyd (1995), quando considera que um domínio científico é um recurso para constituir teoricamente objetos de investigação, assim como "para incorporar e fazer justiça à história da ciência em sua acumulação de conhecimento" (Lloyd, 1995, p.25). Com fundamento nas ideias de Lloyd (1995), Bufrem (2013) afirma que a produção científica desenvolvida em determinados contextos sociais e momentos históricos reflete as mudanças e contradições desse contexto, tanto em sua organização interna quanto em suas aplicações.

Em relação à pesquisa retratada no campo da OC, partiu-se de um modelo alternativo e multidimensional para a análise e discussão dos artigos selecionados, ressaltando natureza, gêneros e dimensões relativas à pesquisa. Esse modelo proposto por Bufrem (2013) considera os polos da prática metodológica de Bruyne *et al.* (1977) como integrantes do processo de pesquisa e os complementa com as dimensões ética e política do processo de investigação científica.

Problematizou-se, neste estudo, como se configuram as relações entre as opções metodológicas adotadas na literatura de OC e o processo dinâmico de pesquisa sob a ótica multidimensional. Como objetivo geral, analisaram-se as opções adotadas na literatura periódica científica de OC sob a ótica do modelo multidimensional da prática de pesquisa. Para tanto, partiu-se

da construção de um *corpus* de artigos selecionado da Brapci a partir de termos relacionados à área. A seguir, categorizaram-se os artigos de pesquisa sob o ponto de vista de seu objeto e delineamento metodológico, bem como de suas relações teóricas representadas por redes de citação. Identificaram-se, então, as opções metodológicas manifestas nessa literatura, relacionando-as com o modelo multidimensional da prática de pesquisa.

Com o intuito de complementar as informações e análises, dentro do contexto da produção e disseminação científica acerca da temática da OC, considerou-se necessário construir um panorama a respeito de quem são os pesquisadores mais profícuos que trabalham com a temática no país. Para tanto, analisaram-se características das autorias, tais como gênero, formação profissional, vínculo institucional, função desempenhada e perfil de publicação desses pesquisadores, destacando-se seus periódicos preferidos para disseminação do conhecimento e, por fim, a modalidade de autoria mais frequente.

Métodos

Para analisar e relacionar as opções metodológicas do campo de OC no Brasil com as dimensões da pesquisa científica que fundamentam a interpretação aqui realizada, este estudo realizou um recorte de artigos científicos da base de Dados Referenciais de Artigos de Periódicos de Ciência da Informação (Brapci) sobre organização do conhecimento, concernente ao período de dez anos, entre 2003 e 2012, com vistas ao aprofundamento das relações entre os aspectos evidenciados. A Brapci, instrumento para a constituição do *corpus*, é resultado de um projeto de pesquisa acadêmica com o intuito de facilitar a pesquisa de documentos e artigos da área. Atualmente ela disponibiliza referências e

resumos de 9 673 textos, publicados em 37 periódicos brasileiros, impressos e eletrônicos, da área de Ciência da Informação (CI), constituindo-se no mais completo repositório da produção científica periódica brasileira em CI.

Partiu-se de um estudo exploratório na Brapci com os termos descritores “organização do conhecimento”, “terminologia” e “conceito” nos campos título, palavra-chave e resumo, obtendo-se um total de 974 artigos. Para atribuir grau de relevância aos registros recuperados empregou-se uma metodologia de pontos incorporada à Brapci, a qual classifica os registros de acordo com a incidência dos termos de busca nesses campos já mencionados, categorizando-se os artigos conforme a relevância para a temática (Freitas *et al.*, 2010). Foram selecionados os artigos que apresentaram ao menos uma estrela cheia para determinar sua relevância, isto é, que continham o descritor em pelo menos dois campos de busca, do que resultou um universo de 177 artigos. Utilizou-se o *software Endnote* para a organização dos dados coletados, bem como o programa Excel para facilitar a análise dos resultados obtidos.

Em um segundo momento de seleção, foi realizada a leitura dos resumos para descartar pesquisas não pertinentes ao tema em foco, restando um *corpus* de 105 artigos no período delimitado, considerado significativo, tanto pela diversidade dos estudos que nele se concretizam, quanto pela representatividade de pesquisadores distinguidos no campo de atuação da OC.

É importante salientar que como estratégia de pesquisa utilizou-se a contagem completa dos autores. Ela consiste na atribuição de um valor inteiro para cada um dos autores e coautores. Em decorrência dessa opção, o número total de artigos foi artificialmente aumentado de 105 (títulos diferentes) para 186 (número de autorias). As análises feitas com base no número de autorias foram relativas à formação acadêmica, ao vínculo institucional, à função desempenhada, bem como à produção individual. Acredita-se que as situações relacionadas a essas variáveis são passíveis de alteração ao longo do período analisado. Já para as demais variáveis (temática e opção metodológica, dentre outras) utilizou-se o total de artigos distintos. Essa opção deve-se ao fato de que, dentre as possibilidades de mensurar a colaboração e identificar o perfil dos produtores na área de OC, a

contagem completa mostrou-se mais adequada do que a direta, também denominada hierárquica, e a fracionada ou proporcional (Urbizagástegui Alvarado, 2002).

Resultados e Discussão

Os resultados (Tabela 1) mostram inicialmente que os periódicos mais devotados ao tema OC são DatagramaZero, com 21,90% dos artigos publicados; Perspectivas em Ciência da Informação, com 12,38%; Ciência da Informação e Encontros Bibli, cada um com 11,43% do total de 105 artigos publicados durante o período.

O primeiro lugar no *ranking*, como veículo de disseminação de trabalhos sobre a OC, obtido pelo periódico DataGramaZero pode ter como justificativa a publicação de um número especial sobre a temática e assuntos circunscritos a ela no ano de 2006. Também foi possível verificar que a maior parte dos artigos foram publicados em revistas pertencentes ao estrato B1 (61,90%), sendo que nenhuma produção foi observada em periódicos classificados como A2 na subárea Ciência da Informação, de acordo com o Qualis/Capes.

Dos 114 autores distintos, somente 32,46% são do sexo masculino, resultado que não diverge de outros estudos de sexo da área de Ciência da Informação, que dão conta de que a maioria dos autores ainda é do sexo feminino. No *corpus* delimitado apenas um autor não pôde ser identificado (0,88%), pois somente suas iniciais constavam no artigo e, por não ter nacionalidade brasileira, tampouco faz parte da Plataforma Lattes.

Em relação ao índice de transitoriedade, verificou-se que 84 das 186 autorias publicaram uma única vez no período, resultando em 45,16% de autores transeuntes. Esse resultado sugere não só uma tendência de consolidação da temática no país, como também indica maior representatividade de alguns que despontam como a elite produtiva sobre a OC.

Quando se aplicou aos 114 autores distintos do *corpus* a lei do elitismo de Price, que postula que a raiz quadrada do total de autores em determinada área se constitui como elite, obteve-se o número de 10,68 autores situados no topo da lista de produção dentro do tema Organização do Conhecimento (Tabela 2). Entretanto, o

Tabela 1. Ranking dos periódicos mais profícuos na temática de Organização do Conhecimento (2003-2012).

Periódico	Qualis	Contribuição	
		Frequência Absoluta	Frequência Relativa %
DataGramZero	B1	23	21,90
Perspectivas em Ciência da Informação	A1	13	12,38
Ciência da Informação	B1	12	11,43
Encontros Bibli	B1	12	11,43
Informação & Informação	B1	8	7,62
Transinformação	A1	8	7,62
Informação & Sociedade	A1	5	4,76
Em Questão	B1	3	2,86
Revista Digital de Biblioteconomia e Ciência da Informação	B1	3	2,86
Biblos	B3	2	1,90
InCID	B1	2	1,90
Ponto de Acesso	B1	2	1,90
Revista Brasileira de Biblioteconomia e Documentação	B1	2	1,90
Revista Ibero-Americana de Ciência da Informação	B1	2	1,90
Tendências da Pesquisa Brasileira em Ciência da Informação	B1	2	1,90
ACB	B2	1	0,95
Arquivística.net	B4	1	0,95
Educação Temática Digital	B4	1	0,95
Informação e Cognição	B5	1	0,95
Liinc em Revista	B1	1	0,95
Perspectivas em Gestão & Conhecimento	B1	1	0,95
Total		105	100,00

Fonte: Dados da pesquisa 2003-2012. Elaboração própria, 2014.

Tabela 2. Autores mais produtivos na área de Organização do Conhecimento (2003-2012).

Autor	Frequência Absoluta	Frequência Relativa %
MONTEIRO, Silvana Drumond	7	3,76
BUFREM, Leilah Santiago	6	3,23
VAN DER LAAN, Regina	6	3,23
CAMPOS, Maria Luiza de Almeida	5	2,69
KOBASHI, Nair Yumiko	5	2,69
LARA, Marilda Lopes Ginez de	5	2,69
TÁLAMO, Maria de Fátima G. Moreira	5	2,69
GOMES, Hagar Espanha	4	2,15
GUIMARÃES, José Augusto Chaves	4	2,15
BRÄSCHER, Marisa	4	2,15
CAFÉ, Lígia	4	2,15
Total	55	29,57

Fonte: Dados da pesquisa. Elaboração própria, 2014.

Em relação à formação e à função desempenhada, não surpreende constatar que 56,45% dos autores possuem título de doutor e que 63,44% mencionaram a docência como principal atividade desempenhada. No que concerne ao vínculo institucional, a Universidade de São Paulo (USP) desponta como a primeira colocada com 14,52%, seguida pela Universidade Federal do Rio Grande do Sul (UFRGS) com 10,22% e pela Universidade Estadual Paulista com 8,06% das autorias (186).

No intento de melhor compreender esses percentuais, buscaram-se os grupos cadastrados e certificados no último censo de 2010 no Diretório de Grupos de Pesquisa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (DGP/CNPq), utilizando-se as seguintes palavras-chave: organização do conhecimento e organização da informação. Foi possível constatar que, dos 90 grupos revocados, com base nessa estratégia de busca, apenas quatro mencionavam vínculo com a USP, sendo que somente um era da área de Ciência da Informação (Grupo Temma). A partir desse resultado, foi possível identificar que, dos 11 autores componentes da elite, quatro são membros do referido grupo.

recorte utilizado expandiu esse valor para 11 autores, objetivando equalizar o ponto de corte em quatro artigos. Observa-se também que a produção desses autores (29,57%) não ultrapassa os 50,00% preconizados por Price para garantir uma elite produtiva.

Algumas inferências podem ser feitas com base nesses dados relacionais, especialmente cotejando-os com outros resultados de pesquisas sobre a área de CI, que mostram uma produção individual ou individualizada. A diferença básica entre esses termos está no fato de que o primeiro se refere à modalidade de autoria, e o segundo, ao processo incipiente de fortalecimento, visibilidade e produção dos grupos de pesquisa. No que concerne à modalidade de autoria dos 105 artigos publicados, observou-se que a maioria foi apresentada na modalidade individual (45,71%), reforçando estudos anteriores na área de CI (Mueller & Pecegueiro, 2001; Vilan Filho, 2010; Nascimento & Gomes, 2012).

Levou-se em conta, para esta pesquisa, que os fundamentos teóricos de autores relacionam-se com suas posições epistemológicas e políticas, concretizadas nos modos de fazer pesquisa e, por decorrência, nas opções metodológicas eleitas para a realização de suas pesquisas e em suas preferências temáticas. Com esse pressuposto, analisaram-se os tipos de pesquisa quanto aos fins, meios, enfoques e tipos de análises e técnicas eleitos por esses autores no domínio da OC. Para complementar essas posturas metodológicas e também epistemológicas, elencaram-se autores utilizados como suporte teórico pelos autores mais produtores do *corpus* analisado.

Realizou-se um estudo das referências dos 55 artigos publicados pelo grupo destacado nesta pesquisa para identificar quais autores representam os fundamentos teóricos dos trabalhos analisados. A autora mais citada foi Ingetraut Dahlberg, com 45 incidências (Figura 1). Esse resultado é condizente com a importância das contribuições da autora para a área de Organização do Conhecimento, em especial com a publicação de seu artigo intitulado "Teoria do conceito", de 1978, na revista brasileira *Ciência da Informação*. Com esse artigo, Dahlberg (1978) estabelece bases teóricas para o entendimento do conceito como estrutura formada por características (enunciados) dos objetos. Essa concepção fundamenta as análises relacionadas à área de Organização do Conhecimento, bem como pode ser relacionada a outros autores basilares, conforme os resultados a seguir.

Ressalta-se que três (30%) dos autores mais citados pelos mais produtivos são brasileiros, sendo que

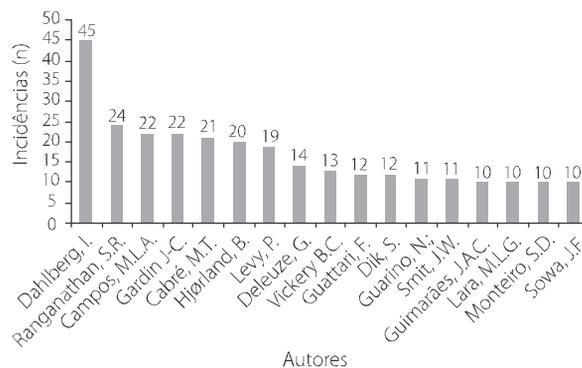


Figura 1. Autores mais citados pelos mais produtivos.

Fonte: Elaboração própria, 2014.

alguns deles também figuram entre os autores mais representativos no *corpus*. O fato sugere a possibilidade de autocitação, embora o que se tenha verificado é que somente 8,5% das citações se caracterizaram como tal.

Entre os autores citados concomitantemente pelos mais representativos do *corpus* (Figura 2), observa-se uma incidência expressiva de Dahlberg com Campos e Dahlberg com Hjørland.

Verificou-se que autores basilares da área de OC figuram entre os mais citados e, em alguns casos, como os mais produtores também. Nesse sentido, uma análise possível seria em relação à presença de artigos voltados à edificação de conceitos teóricos que buscam estabelecer fundamentos do campo.

No que concerne às temáticas consideradas, elas apresentam grande diversidade no que diz respeito às subáreas da Organização e Tratamento da Informação. Sendo assim, optou-se pela descrição somente das que figuram com duas ou mais ocorrências no *corpus* (Tabela 3).

Ao analisar os dados acima, inferiu-se que a maioria dos trabalhos (n=105) abordou temáticas não recorrentes (58,10%), ou seja, esses temas foram trabalhados apenas uma única vez ao longo do período analisado. Dentre as temáticas mais frequentes, a primeira posição é ocupada pela linguagem documentária, e a segunda posição pela terminologia, áreas com presença em atividades diversas relacionadas com a sistematização, representação e edificação de estruturas conceituais que ponderam a lógica e a linguística para seu aprofunda-

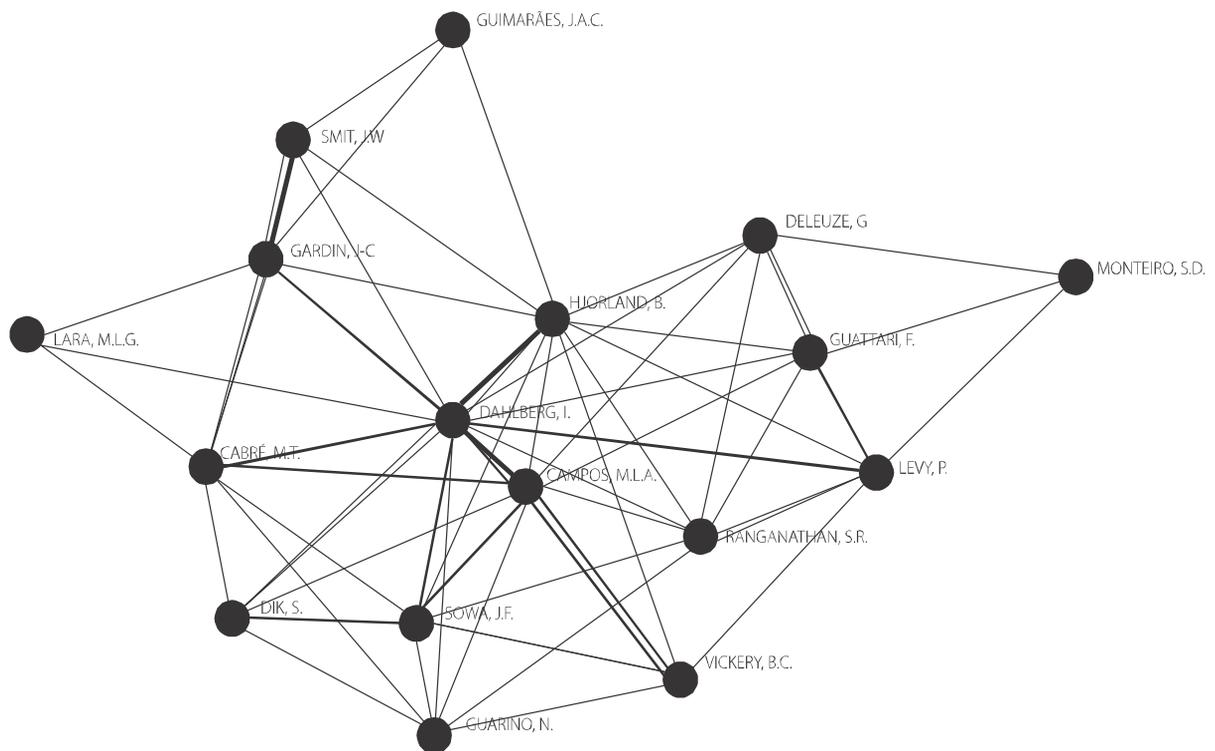


Figura 2. Rede de cocitação dos autores mais representativos do corpus.

Fonte: Elaboração própria, 2014, utilizando o software UCINET.

Tabela 3. Temáticas mais frequentes.

Temática	Frequência	
	Absoluta	Relativa %
Linguagem Documentária	16	36,36
Terminologia	11	25,00
Conceito	8	18,18
Organização do Conhecimento	4	9,09
Organização Virtual do Conhecimento	3	6,82
Representação do Conhecimento	2	4,55
Total	44	100,00

Fonte: Dados da pesquisa. Elaboração própria, 2014.

resultado corrobora a adequação teórica dos trabalhos, fato que justifica a presença das autoras Ingetraut Dahlberg (45 incidências) e Maria Tereza Cabré (25) dentre as mais citadas.

As correntes teóricas relacionadas ao Tratamento Temático da Informação, podem ser classificadas, conforme Guimarães (2008) em: catalogação de assunto (*subject cataloguing*), de influência norte-americana; indexação (*indexing*), de influência inglesa; e análise documentária (*analyse documentaire*), de influência francesa.

Segundo Guimarães *et al.* (2012, *online*), a análise documental tem como orientação principal a corrente francesa, cujo “[...] foco centra-se no processo de tratamento temático da Informação em si, [...] para a identificação e seleção de conceitos e posterior representação e geração de produtos”. Segundo os autores, esse tipo de estudo revela “[...] forte interface com a Linguística e a Lógica”, com especial relevo para a interface da documentação com a linguística (Gardin, 1973) e as questões

relacionadas às linguagens documentárias (Coyaud, 1966).

Ao identificar as pesquisas em relação aos fins, constata a predominância de pesquisas de caráter exploratório (57,14%), o que pode indicar não só que a área tem buscado aplicabilidade empírica de suas teorias já sedimentadas, por meio de estudos de caso específicos de organização de informação e conhecimento em unidades de informação organizacionais e em instituições de diferentes tipos, como também aponta para estudos que visam ampliar a familiaridade com os temas propostos.

Em relação aos 34 trabalhos (32,38%) categorizados como descritivos, salienta-se que esse tipo de estudo se caracteriza por um aprofundamento teórico mais intenso em relação ao objeto estudado, distinguindo-se do que vem sendo observado nos estudos exploratórios, mais incipientes por ensejarem um conhecimento aproximativo das características e aspectos de um objeto de pesquisa. Por sua vez, os estudos descritivos podem representar um amadurecimento na pesquisa teórica da área, por meio do estabelecimento de relações específicas entre variáveis que envolvem as pesquisas. Já os estudos metodológicos, mais complexos, com nove incidências no *corpus* analisado, geralmente configuram-se em propostas de criação de modelos, metodologias e estruturas para a organização da informação em contextos determinados. Em alguns casos, os estudos metodológicos remetem à necessidade de conhecimento sobre tecnologias da informação e comunicação, quando não são exclusivamente modelos conceituais. Nesse caso, exigem-se estudos piloto para validação das propostas. Um exemplo de estudo metodológico no *corpus* relata modos de estruturação da Base Brapci na *Web* para acesso universal.

Sobre a última modalidade de pesquisa quanto aos fins, encontram-se os estudos avaliativos, com duas incidências, cujo foco pode ser um objeto, serviço, fluxo, unidade de informação apresentando critérios ou parâmetros a partir de princípios de avaliação. Essas duas últimas tipologias não são expressivas por sua incidência, devido à complexidade de seus processos.

Pode-se verificar que duas das matrizes teóricas relativas às interdisciplinaridades em Análise Docu-

mental, postuladas por Guimarães (2010, *online*) no artigo "Análise documental: concepções do universo acadêmico brasileiro em Ciência da Informação" correspondem à Matriz Quantitativa e à Matriz Tecnológica, que englobam temas relativos à Ciência da Computação e à Informática. Segundo o autor, essas figuram entre as matrizes menos citadas pelos pesquisadores. Entretanto, o "elemento inesperado está na constatação de que a maioria dos pesquisadores relacionou os objetivos da análise documental à Recuperação da Informação, ação que depende diretamente do funcionamento adequado das tecnologias, bem como dos cálculos quantitativos". Esse aspecto, segundo o autor, contradiz a tônica das respostas da pesquisa por ele realizada, e "vai de encontro também com a literatura da área, particularmente nos anos 70, que tradicionalmente se valeu dos modelos matemáticos para o processo de recuperação da informação" afirma Guimarães (2010, *online*).

As pesquisas bibliográficas são características da área de OC e visam estimular seu desenvolvimento conceitual. A predominância desse tipo de pesquisa (76,19%) pode reforçar o pensamento de que a institucionalização de qualquer área do conhecimento ocorre mediante a consolidação teórica pela comunidade científica, com o estabelecimento de conceitos, técnicas e metodologias próprias do campo em questão.

Em relação aos enfoques, a maior incidência observada foi o enfoque teórico (49,52%), seguido pelo enfoque linguístico (17,14%), corroborando o exposto acima, já que a maioria dos trabalhos caracterizados como pesquisas bibliográficas apresenta enfoque teórico.

Entre alguns dos autores mais citados, isto é, com mais de três publicações no período estabelecido para esta pesquisa, encontra-se com especial ênfase a preferência ao enfoque teórico, abrangendo estudos relacionados à Filosofia da Linguagem, à Teoria do Conceito, à Teoria da Classificação e à Epistemologia, entre outras correntes filosóficas. O enfoque linguístico engloba temas como Semiologia, Análise do Discurso, Semântica, Linguística Textual, Terminologia, Semiótica, Linguística Documentária, Lógica, Análise de Conteúdo, entre outros, coincidindo com os resultados do estudo de Guimarães (2010, *online*). Com menor frequência (6,67%), aparece o enfoque histórico, que considera a

construção histórica da ciência para a realização de análises.

Em relação ao tipo de análise, a mais frequente foi a análise documental, com 42,67% dos artigos, seguida pela análise de conteúdo, com 30,67%. Os demais tipos de análise, como a categorial, a de domínio e a de discurso, obtiveram juntos um percentual menos expressivo (14,67%) em comparação aos dois primeiros. Outros tipos de análise apresentaram-se de modo ainda mais pulverizado em 7 categorias.

Conclusão

O presente estudo possibilitou traçar um panorama das publicações na área de OC, principalmente no que concerne às características estruturais (opções metodológicas) e estruturantes (autores e autorias) que compuseram o cenário das pesquisas científicas no Brasil nos últimos dez anos (2003-2012).

A presença de autores brasileiros entre os mais citados, mesmo com um percentual de 30%, reforça a ideia de que comunidade científica brasileira tem buscado articular-se com a comunidade internacional e com autores basilaes, acompanhando o desenvolvimento e a evolução do domínio da OC. Sem dúvida, esse fato não prescinde da inquestionável importância dos autores internacionais para sua constituição teórica enquanto área de pesquisa, assim como das influências das distintas correntes de Organização e Tratamento Temático da Informação para o desenvolvimento e enriquecimento da OC em âmbito nacional. Entretanto, não se pode deixar de destacar o amadurecimento dos pesquisadores nacionais e a constituição de laços colaborativos entre eles.

A realidade observada depõe em favor da construção de um *ethos* próprio da OC no país, pois é inquestionável que a institucionalização de qualquer área do conhecimento ocorre mediante sua consolidação teórica pela comunidade científica, com o estabelecimento de conceitos, técnicas e metodologias que a distingam das demais. Essa construção se caracteriza por ser uma edificação realizada pelos e entre os agentes envolvidos, os quais estabelecem relações, práticas e preceitos ético-políticos identificados como exclusivos de uma comunidade. Cumpre ressaltar que aqui se considera o conceito de comunidade como revelador de uma identidade individualizante e distintiva, mas aglutinadora. Isso significa que, para pertencer ou transitar em determinada área, é imprescindível que se partilhe desse *ethos* que se constitui das e nas *práxis* empreendidas pelos membros que privam não só desse código e objetivo comum, mas também buscam o reconhecimento e a expansão da área perante a comunidade científica.

Por fim, ressalta-se que uma futura pesquisa, que estabeleça as relações de citação, bem como identifique a colaboração intra e interinstitucional, presentes no *corpus* analisado, pode esclarecer ainda mais de que maneira a área tem se sedimentado ao longo dos anos. Além disso, será possível verificar sob quais aspectos epistemológicos a OC se edifica e se torna reconhecida pela Ciência da Informação.

Agradecimentos

Trabalho apoiado por concessão de bolsa da Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp).

Referências

Bruyne, P. et al. *Dinâmica da pesquisa em ciência sociais: os polos da prática metodológica*. 2.ed. Rio de Janeiro: F. Alves, 1977.

Bufrem, L.S. Configurações da pesquisa em ciência da informação. *DataGramaZero: Revista de Informação*, v.14, n.6, 2013. Disponível em: <http://www.dgz.org.br/ago14/F_1_aut.htm>. Acesso em: 9 ago. 2014.

Coyaud, M. *Introduction à l'étude des langages documentaires*. Paris: Klincksieck, 1966.

Dahlberg, I. A referent-oriented analytical concept theory of interconcept. *International Classification*, v.5, n.2, p.142-150, 1978.

Freitas, J.L. et al. Proposta de metodologia para a recuperação da produção científica em ciência da informação na base Brapci. *Ponto de Acesso*, v.4, n.3, p.45-67, 2010.

Gardin, J.-C. Document analysis and linguistic theory. *Journal of Documentation*, v.29, n.2, p.137-168, 1973.

Guimarães, J.A.C. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). *Revista Ibero-americana de Ciência da Informação*, v.1, n.1, p.77-99, 2008.

Guimarães, J.A.C. Análise documental: concepções do universo acadêmico brasileiro em ciência da informação. *DataGramaZero*, v.1, n.1, 2010. Disponível em: <http://www.dgz.org.br/fev10/Art_02.htm>. Acesso em: 9 ago. 2014.

Guimarães, J.A.C. *et al.* A dimensão interdisciplinar da análise documental nos contextos brasileiro e espanhol no âmbito da organização do conhecimento. *DataGramaZero*, v.13, n.6, 2012. Disponível em: <http://www.dgz.org.br/dez12/Art_07.htm>. Acesso em: 9 ago. 2014.

Lloyd, C. *As estruturas da história*. Rio de Janeiro, Zahar, 1995.

Mueller, S.; Pecegueiro, C.M.P.A. O periódico ciência da informação na década de 90: um retrato da área refletido em seus artigos. *Ciência da Informação*, v.30, n.2, p.47-63, 2001.

Nascimento, B.S.; Gomes, M.Y.F.S.F. A ciência da informação no Brasil: um retrato da área através do estudo de autoria. In: Encontro Nacional de Pesquisa em Ciência da Informação, 13., 2012, Rio de Janeiro. *Anais eletrônicos...*, Rio de Janeiro: Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação, 2012. p.1-21. Disponível em: <<http://www.eventos.econgressos.com.br/metodo/enancib2012/arearestrita/pdfs/19487.pdf>>. Acesso em: 9 ago. 2014.

Urbizagástegui Alvarado, R. A lei de Lotka na bibliometria brasileira. *Ciência da Informação*, v.31, n.2, p.14-20, 2002.

Vilan Filho, J. *Autoria múltipla em artigos de periódicos científicos das áreas de informação no Brasil*. 2010. Tese (Doutorado em Ciência da Informação) - Departamento de Ciência da Informação e Documentação, Universidade Federal de Brasília, Brasília, 2010.

Marketing en el campo de las disciplinas informativas: tendencias de investigación en América Latina (1985-2012)¹

Marketing in the field of information disciplines: Research trends in Latin America (1985-2012)

Carlos Luis GONZÁLEZ-VALIENTE²

Resumen

Se esclarecen las tendencias de investigación relativas a la mercadotecnia en el campo de las disciplinas informativas, a partir de la producción científica de América Latina. Se utiliza como fuente de consulta las revistas electrónicas indizadas por el *Directorio Latindex*. Se recuperaron 108 registros bibliográficos a los cuales se les aplicaron los indicadores bibliométricos de productividad por años, revistas, autores, idioma y análisis de coocurrencia de términos. Se evidencia que el mayor índice de productividad radica en el contexto de las revistas, idioma y autores brasileños. Tras el análisis de coocurrencia se conformaron cuatro clúster principales que denotaron como tópicos más tratados: la mercadotecnia del profesional de la información en el mercado de trabajo, la mercadotecnia en bibliotecas, la mercadotecnia de productos y servicios de información, y la gestión de información de mercadotecnia. Desde otro punto de vista se identificaron tendencias a partir de cuatro rasgos en las investigaciones, tales como: el objeto mercadotécnico, el contexto organizacional, la perspectiva disciplinar y las técnicas y herramientas de marketing más exploradas. Se comprobó que el marketing en las Ciencias de la Información es un tópico que no ha estado ajeno a la investigación en la región latinoamericana, además se logró caracterizar la productividad de un grupo de revistas (51.4%) que no son indizadas por bases de datos de la corriente principal.

Palabras clave: América Latina. Disciplinas informativas. Indicadores bibliométricos. Latindex. *Marketing*. Produção científica. Tendencias de investigación.

Abstract

This study seeks to clarify research trends related to marketing in the field of information disciplines from Latin American scientific production. As a reference source, we used the electronic journals indexed in the Latindex Directory. Some bibliometric indicators were applied to 108 documents; these indicators were as follows: productivity by year, journals, authors, language, and finally term co-occurrence analysis. It was evident that the highest rate of productivity was found in Brazilian journals and authors, and in the Portuguese language. The four main clusters in term co-occurrence analysis in the most relevant topics were as follows: marketing of the information professional in the labor market, library marketing, marketing of information products and services, and marketing information management. From another point of view, four kinds of traits were also identified, as follows: the marketing object, organizational context, disciplinary perspective, and the most commonly used marketing tools and techniques. It was found that marketing in Information Sciences is a relevant topic in the Latin American research activity. In addition, it was possible to establish the productivity of some journals (51.4%) that are not included in the most relevant databases.

Keywords: Latin America. Information disciplines. Bibliometric indicators. Latindex. *Marketing*. Scientific production. Research trends.

¹ Trabajo presentado en el VII Seminario Internacional sobre Estudios Cuantitativos y Cualitativos de la Ciencia y la Tecnología "Prof. Gilberto Sotolongo Aguilar" en XIII Congreso Internacional de Información - INFO'2014. Habana, Cuba.

² Grupo Empresarial de la Industria Sidero Mecánica, Departamento de Informática y Gestión de la Información. Carretera Cujae, km 2½, La Habana, Cuba. E-mail: <carlos.valiente@fcom.uh.cu>.

Recibido el día 6/6/2014 y aceptado para su publicación el 12/9/2014.

Introducción

Las disciplinas informativas son consideradas por Guinchat *et al.* (1990) como el campo en el cual se ubican tres grandes grupos: la archivística, la bibliotecología y la documentación. Sin embargo, Pérez Matos (2011, p.4) las concibe como “aquellas que aparecen enmarcadas en determinadas instituciones, - bibliotecología y archivística- y las que mantienen un carácter institucional completamente independiente - bibliografía y ciencia de la información- y que tienen un objeto de estudio en común: la información”. Aunque Zins (2007a) destaca que, paralelo a la información, se encuentran el dato y el conocimiento como importantes categorías de su estudio. La nomenclatura genérica de los grupos que integran tales disciplinas varía según el contexto geográfico; pero, en este artículo se les reconocerán como el amplio dominio de las Ciencias de la Información.

Las relaciones interdisciplinarias que estas Ciencias de la Información (CI) mantienen con otras áreas del conocimiento han precisado de la construcción de un cuerpo teórico y práctico que está dotado de múltiples enfoques, perspectivas y corrientes de pensamiento. Tal es el caso de la mercadotecnia, la cual, proviniendo de las Ciencias Administrativas, Económicas y Contables; ha sido ubicada dentro de los veintiocho esquemas de clasificación del mapa del conocimiento de las CI que integra Zins (2007b) en los siguientes niveles: 7.3.1, como parte de las *Áreas de aplicación* que define Alan Gilchrist; 5.4, como parte de la *Industria de la información* que Donald Hawkins propone; 4.4, como parte del *Mercado de información* estructurado por Haidar Moukdad y 3.7, como parte de las *Disciplinas gerenciales* que concibe Lena Vania Pinherio.

De forma explícita, González-Valiente *et al.* (2014, *online*) declaran que las relaciones entre la mercadotecnia y las CI “se ha dado en la manera en que los elementos de mercadeo pueden ser aplicados a los productos y servicios de información que se generan en las organizaciones”. Esta idea constituyó la base para que previamente Gupta (2006), en un estudio bibliométrico, indicara las fuentes que han publicado artículos sobre la mercadotecnia de tales productos y servicios informativos. En su investigación, partiendo del criterio

de que es abundante lo que se ha publicado sobre esa área del conocimiento en el campo informacional, fue declarada la importancia de investigar su crecimiento cronológico y temático; así como su comportamiento según las distintas áreas geográficas.

En el contexto específico de Latinoamérica, no se han encontrado exhaustivas y sólidas referencias de materiales publicados donde se hayan caracterizado tales investigaciones producidas en la región. Es por ello que el objetivo de este artículo es el de esclarecer las tendencias de investigación relativas a la mercadotecnia en el campo de las disciplinas informativas, detectadas desde la producción científica de América Latina. Para ello se explorarán indagaciones que se hayan orientado a identificar tendencias similares. Además, serán determinados algunos indicadores bibliométricos que permitan identificar las respectivas tendencias.

Amaral (1996), concentrándose en el dominio brasileño, examinó lo que en ese país se había producido de *marketing* entre 1972 y 1994. Sus principales fuentes de consulta fueron libros, capítulos de libros, tesis de disertación, monografías y artículos de periódicos. Desde ahí la autora identificó 97 referencias, distribuidas por un total de 45 autores. Mientras que Ponjuán (1999), mediante una búsqueda en la base de datos *Library and Information Science Abstract (LISA)*, mostró las tendencias de investigación asociadas a los diferentes elementos de la Gestión. En sus resultados (a partir de 2.632 referencias obtenidas) se evidenció que la mercadotecnia había sido una de las temáticas más tratadas en los últimos treinta años, con cifras superiores a las de otros temas como *gestión de recursos de información, gestión de archivos y registros administrativos, gestión de proyectos, gestión de la calidad total*, y demás. Por otra parte Amaral (2000), tomando como referencia las fuentes LISA y *Library Literature*, exploró la aparición de las cuatro P's del mercadeo en unidades de información durante los años 1975-1995. Aquí fue corroborada la existencia de limitaciones sobre la existencia de una terminología específica sobre *mercadotecnia de información* en el área de las CI. Paralelamente, Gupta y Jain (2009) encaminaron un análisis de la literatura periódica en la base de datos LISA, en donde estudiaron la productividad sobre esta temática, pero solo de las revistas procedentes de la India (cobertura: 1980-2003). Estos autores identificaron

patrones de autoría, colaboración, crecimiento y tópicos emergentes a partir de 125 artículos procedentes de 38 revistas.

Entre los trabajos anteriormente citados se percibe cómo autores latinoamericanos, como Ponjuán y Amaral, han investigado bibliométricamente el fenómeno de la mercadotecnia; pero orientan sus análisis sobre literatura indizada en bases de datos que no son de la región latinoamericana. En otros casos, los estudios solo se han reducido a contextos muy limitados, abarcando periodos poco emergentes para los momentos actuales. Estas prácticas, a instancias de Rodríguez-Gallardo (2009), sesgan los resultados obtenidos porque no son proporcionados los datos reales de la producción científica del área.

Métodos

Para la consulta de materiales bibliográficos que han abordado el tema en cuestión, se utilizó como método de investigación el análisis documental clásico. Mientras que la identificación y esclarecimiento de las tendencias se hizo a partir de los principios de la bibliometría, debido a su relevancia en la caracterización de los variados perfiles científicos (Miguel *et al.*, 2006). Concerniente al hallazgo de los registros se utilizó como fuente de consulta el *Directorio* del sistema de información *Latindex* <<http://www.latindex.unam.mx/latindex/directorio.html>>, el cual indiza revistas científicas clasificadas por disciplinas, no solo de América Latina y el Caribe, sino también de España y Portugal; países que también fueron usados para hacer más provechoso el análisis y la discusión de los resultados. *Latindex* no solo fue seleccionada por su cobertura geográfica y temática, sino también por presentar una base de datos bien actualizada (Verdecia-Gutierrez, 2010) y por estar regida por adecuadas normas, criterios e indicadores de calidad (Cetto, 1998). Es importante destacar que aunque no es una práctica muy común, debido a las particularidades de su infraestructura, *Latindex* se ha convertido en objeto de consulta para el desarrollo de estudios bibliométricos en varias temáticas ubicadas tanto dentro como fuera del dominio de las CI (Alcain & Ascurain, 2002; Sieres, 2006; Rodríguez-Gallardo, 2009).

En el Directorio fue escogida la disciplina *Bibliotecología*, la cual contiene 159 revistas. Sobre las disciplinas informativas el *Directorio Latindex* expone, independientemente, el campo Bibliotecología por un lado y Ciencia de la Información por otro. Se escogió el de Bibliotecología por poseer más títulos (159) y porque además incluye todas las revistas listadas en el campo Ciencia de la Información (146). La búsqueda fue ejecutada sobre aquellas que permitieron el acceso electrónico a sus materiales mediante la opción *Enlace de Revistas Electrónicas*, entre las que solo figuraron 57. Este *Enlace* localiza automáticamente las publicaciones incluidas en el Directorio que tienen versión en línea e informa sobre el tipo de acceso, los formatos en que se presenta y la cobertura temporal, estableciendo un enlace con la dirección electrónica de la revista (Flores *et al.*, 2009). Los términos usados en las ecuaciones de búsqueda fueron *mercadotecnia* y sus posibles familias de palabras como *mercadeo* y *mercado*, en las variantes idiomáticas de español, inglés y portugués. No se definieron límites mínimos en cuanto a cobertura temporal, pero sí un máximo hasta el 2012. Se le dio prioridad a los registros que destacaron tales términos en el título y las palabras clave.

Para el esclarecimiento de las tendencias se emplearon algunos indicadores bibliométricos que permitieron caracterizar el comportamiento de la temática. Se usaron algunos de los que Schneider (2004) propone dentro del grupo de *análisis de publicaciones*, tales como: Productividad por revistas; Productividad por años; Productividad autorial; Productividad idiomática; Análisis de coocurrencia de términos.

Se realizó el cálculo de los cuatro primeros indicadores, y en el caso del análisis de coocurrencia se procesaron aquellos términos que coocurrieron dos o más de dos veces. Ello permitió visualizar las frecuencias y relaciones que presentaron los descriptores. Los términos que se encontraban en portugués e inglés fueron traducidos al español para el desarrollo de dicha coocurrencia. En un primer momento se conformó una biblioteca a partir del gestor bibliográfico *EndNote X5*, el cual facilitó la elaboración de una base de datos con las referencias obtenidas y así se normalizaron las entradas de cada registro. Posteriormente fue usado el *software Bibexel* para el análisis de la coocurrencia que se efectuó

a partir de los descriptores extraídos del título y las palabras clave de los artículos. Por último, se empleó el programa *VOSviewer* 1.4.0 para la generación y visualización de los mapas que se elaboraron a partir de un fichero *net* creado en Bibexel.

Resultados y Discusión

Productividad temática por años y revistas

Durante el ejercicio de búsqueda se obtuvieron 108 artículos para el periodo 1985-2012; cuya distribución alcanza un límite de 33 revistas a partir del total de las 57 que poseen enlace electrónico a sus materiales. A través de la Figura 1 se evidencia cómo el índice de crecimiento temático ha ido aumentando con el curso del tiempo. Un 73,14% de las investigaciones han sido publicadas durante la década 2000-2012, develándose que por este plazo la productividad ha aumentado casi tres veces más con respecto a lo publicado desde 1985 hasta el 2000 (26,86%). Es detectable, a su vez, que la mayor cantidad de estudios se publicaron entre los años 2006-2009, periodo en el que está concentrado el 30,50% de la productividad. Este incremento puede que se deba a los grandes intereses por expandir el mercado de trabajo de los profesionales

de la información, ya que la mayoría de los estudios presentes en esta fecha se enfocan en este tipo de abordajes.

Para determinar las revistas que más han colaborado se procedió a representar arbitrariamente aquellas que han publicado más de cinco artículos. A través de la Figura 2 se visualizan como muy productivas, las procedentes de Brasil. En primera instancia figuraron *Ciência da informação* (Brasil) y *Ciencias de la Información* (Cuba), aunque en cada caso respectivamente, solo tres de sus artículos se han publicado después del 2000. Ello confirma que actualmente no se comportan como las más prolíferas. Contrariamente; *Biblios* (Perú), *Biblionline* (Brasil), *Informação & Sociedade: Estudos* (Brasil) y *Perspectivas em Ciência da Informação* (Brasil) concentran sus estudios en la última década; considerándose éstas como las más emergentes. Importante es destacar que de las revistas listadas en la (Figura 2), cuatro están siendo indizadas, indistintamente, por las bases de datos *Scopus* y *Web of Science* (WoS); garantizándose así la alta calidad y buena visibilidad de gran parte de las investigaciones (48,14%) sobre el tópico en cuestión.

Enfocando un poco el análisis desde otra arista, es detectable la baja contribución de los países en relación al total de aquellos que pertenecen a la región latinoamericana, ya que las 33 revistas identificadas proceden solamente de ocho países. Éstos son: Brasil

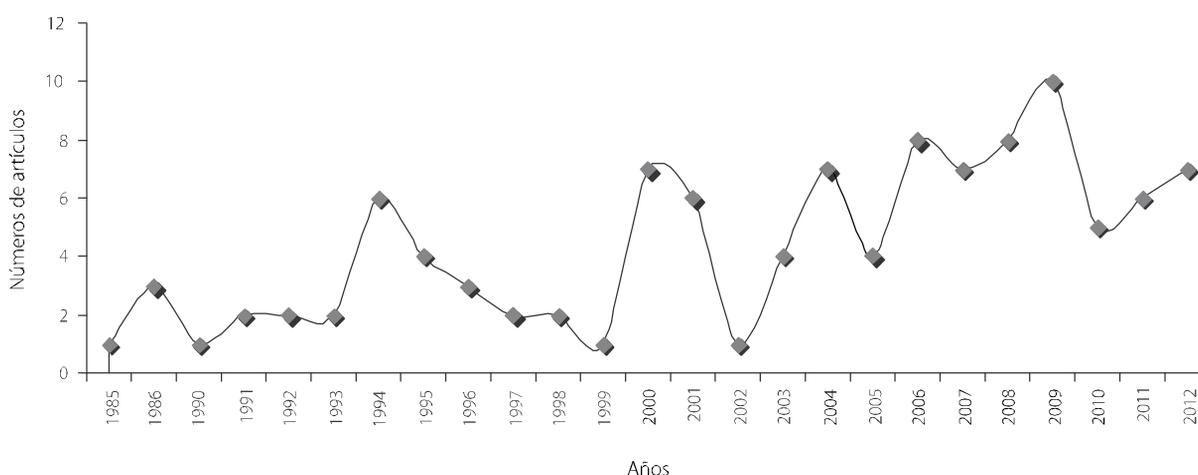


Figura 1. Productividad de las investigaciones por años.

Fuente: Elaboración del autor a partir de los datos obtenidos (2014).



Figura 2. Comportamiento de las revistas más productivas.

Fuente: Elaboración del autor a partir de los datos obtenidos (2014).

(41%), España (25%), Cuba (10%), México (9%), Perú (6%), Argentina (3%), Colombia (3%) y Venezuela (3%). Se mantiene Brasil como líder, región desde la cual 13 revistas han publicado estudios sobre *marketing*; evidenciándose algunos títulos que no aparecen reflejados en la Figura 2, tales como: *Em Questão* (4), *Pesquisa Brasileira em Ciência da Informação e Biblioteconomia* (3), *Transinformação* (3), *Informação & Informação* (3), *Comunicação e Informação* (2), *Revista Digital de Biblioteconomia e Ciência da Informação* (2) y otros cuya productividad es de solo un artículo.

Estos resultados que asientan al dominio brasileño como el más relevante son paralelos con datos actuales que muestra *SCImago Journal & Country Rank*, el cual destaca a Brasil como el país latinoamericano que, en el área de la Bibliotecología y Ciencias de la Información (BCI) para el 2012, mantiene un mayor número de documentos (78) e *índice Hirsch* (17) en comparación con el resto de los países de la región <http://www.scimagojr.com/countryrank.php?area=3300&category=3309®ion=Latin+America&year=2012&order=it&min=0&min_type=it>. Aunque indagaciones muy puntuales, como las de Licea de Arenas *et al.* (2000), Ríos Gómez y Herrero Solana (2005) y Martínez Rodríguez y Solís Cabrera (2013), han demostrado que las revistas brasileñas son las que más publican artículos científicos sobre BCI en toda América Latina.

Productividad idiomática y autoral

Se identificó que el 50,92% de los artículos han sido publicados en español, aunque un 53,00% de las investigaciones concentradas en la última década han sido en portugués. Esto precisa el desplazamiento del portugués por el español en materia de mercadotecnia. Similarmente, los autores más productivos son nativos de esta lengua, y proceden precisamente de Brasil. En primera instancia se destaca Sueli Angelica do Amaral con nueve artículos (9,3%) (Figura 3). Una de las áreas de formación e investigación de esta autora es la mercadotecnia de información, y sobre esta área realizó su tesis de maestría y doctorado en Ciencias de la Información por la Universidad de Brasilia. Actualmente es la líder de investigación en el área de mercadeo dentro del Programa de Postgrado en Ciencia de la Información. Otra autora fecunda ha sido Alzira Karla Araújo da Silva, a quien le pertenecen cinco artículos (4,6%). Su vínculo con este tópico devine gracias a que funge como profesora en la modalidad pregrado de las disciplinas *Mercadotecnia en unidades de información y Mercadotecnia bibliotecaria*, en los cursos de Biblioteconomía y Archivología de la *Universidade Federal da Paraíba*. Por último, con solo cuatro artículos (3,7%), se encuentra Angela María de Oliveira, quien se desenvuelve como bibliotecaria de la *Universidade Estadual de Ponta Grossa* y cuya tesis de maestría fue sobre mercadotecnia en el contexto de las bibliotecas públicas³.

Análisis de coocurrencia de términos

Mediante el análisis de coocurrencia propiciado a través de *VOSviewer* 1.4.0 quedaron conformados nueve clúster principales, de los cuales se harán breves descripciones de cuatro de los más representativos (Figura 4). El clúster 1 contiene un total de trece ítems, siendo *profesional de la información* el de mayor densidad; y manteniendo una mayor fuerza de relación (fr) con las categorías *mercado de trabajo* (fr: 7) *mercadotecnia personal* (fr: 4) y *promoción* (fr: 3). Estas relaciones denotan las exploraciones asociadas a mostrar

³ La información de apoyo que se expone sobre estas autoras ha sido extraída de la *Plataforma Lattes* <<http://lattes.cnpq.br/>>, la cual publica currículum profesionales de investigadores asociados al *Centro Nacional de Desenvolvimento Científico y Tecnológico* (CNPq) de Brasil.

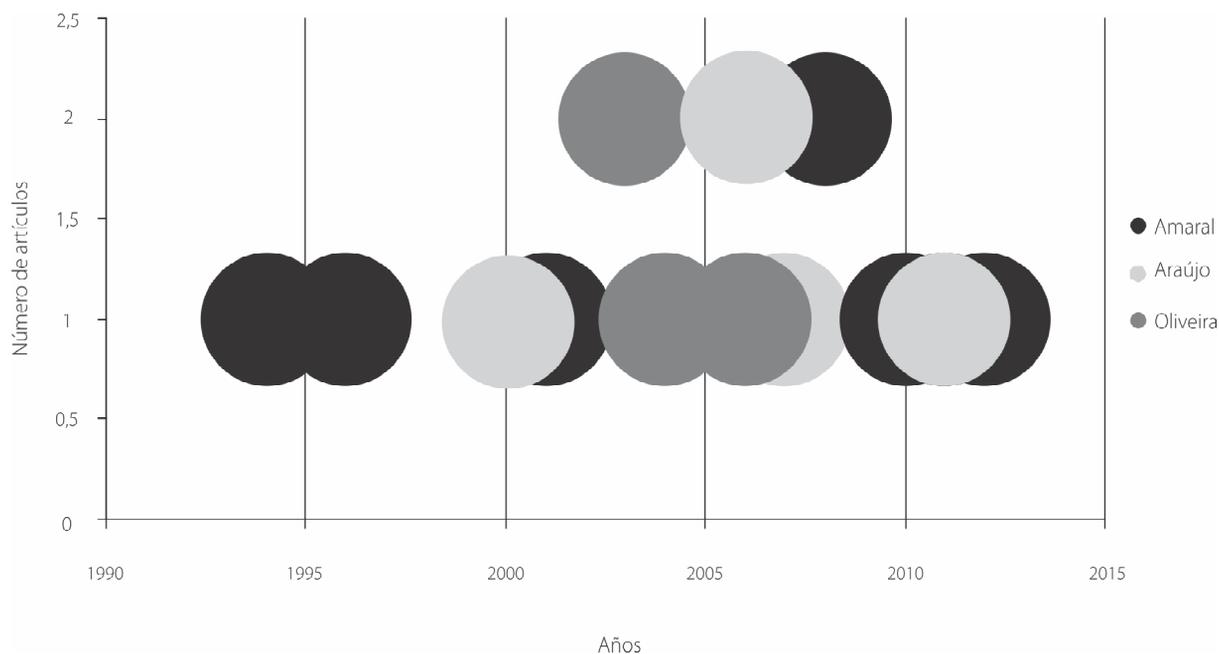


Figura 3. Comportamiento de los autores más productivos según la serie temporal.

Fuente: Elaboración del autor a partir de los datos obtenidos (2014).

la figura de este profesional como un objeto de interés para el escenario laboral. Nótese cómo también dentro de las categorías aquí incluidas están *mercadotecnia profesional, ofertas de empleo e imagen*. Al parecer, esta tendencia parece examinarse mayormente desde la perspectiva de los grupos disciplinares, ya que en este clúster es donde más se concentran los términos que dan nombramiento al dominio informacional, tales como: *Archivología, Biblioteconomía y Documentación*.

Por otra parte el clúster 2, con un total de diez ítems, resalta por agrupar descriptores relacionados con el entorno de las tecnologías; aunque es el clúster que posee menor densidad y mayor aislamiento dentro del mapa. Como categoría más intensa se encuentra *mercadotecnia*, quien tiene una gran fuerza de relación con *bibliotecas* (fr: 18), *mezcla de mercadotecnia* (fr: 8), *gestión del conocimiento* (fr: 7) y *mercadotecnia de información* (fr: 6); aunque aludiendo un poco sobre las tecnologías, las fuerzas más potenciales se dan con *ciberspacio* (fr: 3), *sitios Web* (fr: 3) e *Internet* (fr: 3) respectivamente.

Al clúster 3 están adheridos diez descriptores que responden generalmente a los elementos *productos de*

información y servicios de información; éstos con gran fuerza de relación entre sí (fr: 11). También es detectable la marcada presencia de las categorías *información de mercadotecnia, sistemas de información de mercadotecnia, base de datos y mercado*. Mientras que el cuarto clúster 4 posee diez descriptores al igual que el 3, en donde, *gestión de información* es el más denso. Sus mayores relaciones están dadas con los términos *mercadotecnia* (fr: 8), *gestión del conocimiento* (fr: 4), *mercado* (fr: 3), *servicios* (fr: 3) y *ciencia de la información* (fr: 3). Otros de los ítems distintivos en este clúster lo son *mercado de información, industria de la información y usuarios*.

La coocurrencia también ha servido para indagar sobre cuatro rasgos muy particulares, tales como el objeto mercadotécnico, el contexto organizacional, la perspectiva disciplinar y las técnicas y herramientas de mercadotecnia que más se han tratado. Desde la Figura 5 se visualizan las relaciones de términos, cuestión básica para identificar estos rasgos. Entre el principal objeto de mercadeo está *servicios de información*, con una Frecuencia de Aparición (fa) de 22 veces. Ello confirma una vez más la concepción de que la profesión informativa es, por excelencia, una profesión de servicios. Similarmente es percibido como otro objeto distintivo

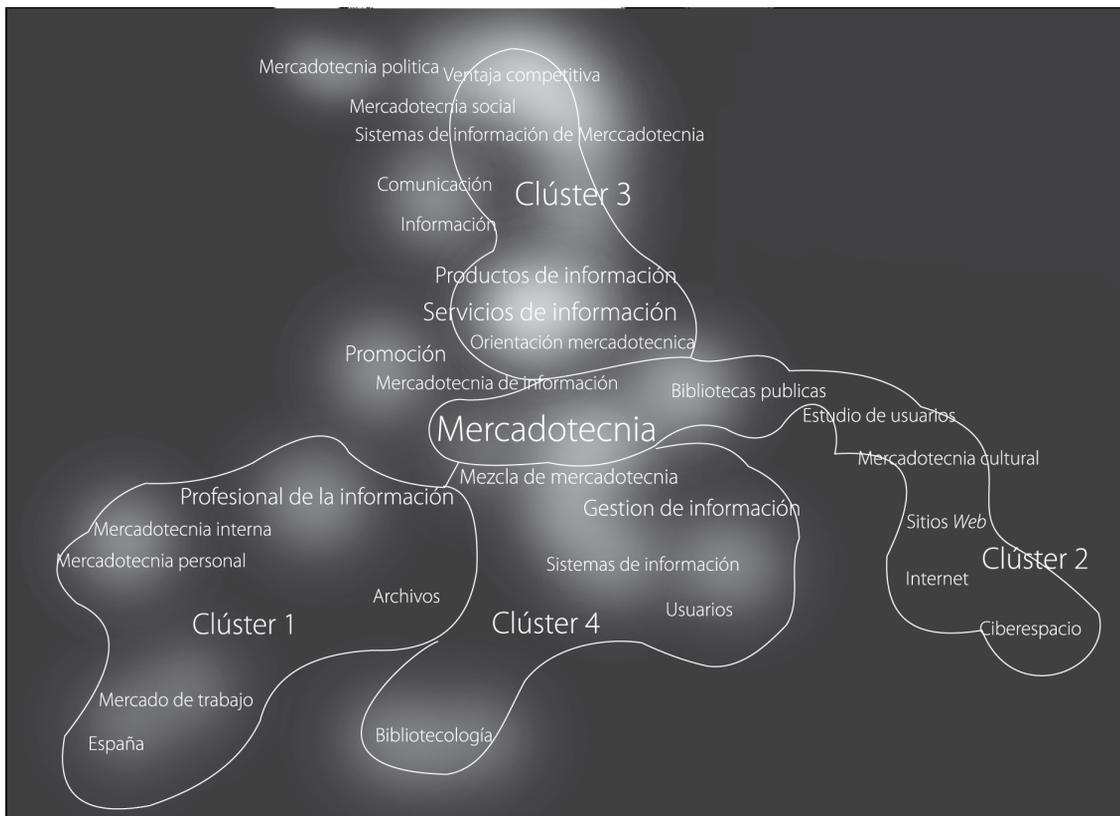


Figura 4. Mapa de densidad de clúster de la coocurrencia de términos.

Fuente: VOSviewer 1.4.0. (2014).

la figura del *profesional de la información* (fa: 18), al cual se le relacionan técnicas precisas de *mercadotecnia personal* y *promoción* en el contexto del *mercado de trabajo*. Cabe afirmar que una de las utilidades del marketing en el campo informacional es la de ofertar a sus profesionales como productos humanos que pueden ser oportunos en el marco de un mercado laboral que está necesitando cada vez más de experticia en el manejo de la información. De ahí que los abordajes del mercadeo desde el ángulo de lo profesional tienen una orientación dual; una interna, dirigiéndose a la adquisición de competencias asociado al marketing como subdisciplina dentro de las CI, y otra externa, enfocada a la aplicabilidad del marketing para dar a conocer las competencias de estos profesionales de manera que salden las exigencias de los mercados laborales actuales. En menor medida ha primado la categoría *productos de información* con una frecuencia de aparición de 13 veces.

Referente a los contextos organizacionales, un primer plano lo ocupan las *unidades de información* (fa: 19), y seguido, las *bibliotecas* (fa: 18). De este último sector, han sido más frecuentes las de tipo *universitarias* (fa: 8), y luego las *públicas* (fa: 7). Oportuno es de resaltar que los *archivos* (fa: 5) están llegando a ser otra de las organizaciones informativo-documentales que muestran acrecentado interés por la orientación al mercado. Mientras que desde la perspectiva disciplinar prima la *Ciencia de la Información* (fa: 5), relacionándose ampliamente con la *Gestión de información* y la *Gestión del conocimiento*. En otra medida se encuentra la *Biblioteconomía* (fa: 4), reflejando gran vínculo con el *mercado de trabajo* (fr: 3). Y en última instancia ha figurado la *Archivología* (fa: 2), siendo un factor importante que no solo se haga alusión a sus respectivas instituciones, sino también a la disciplina a la cual responden. Asociado

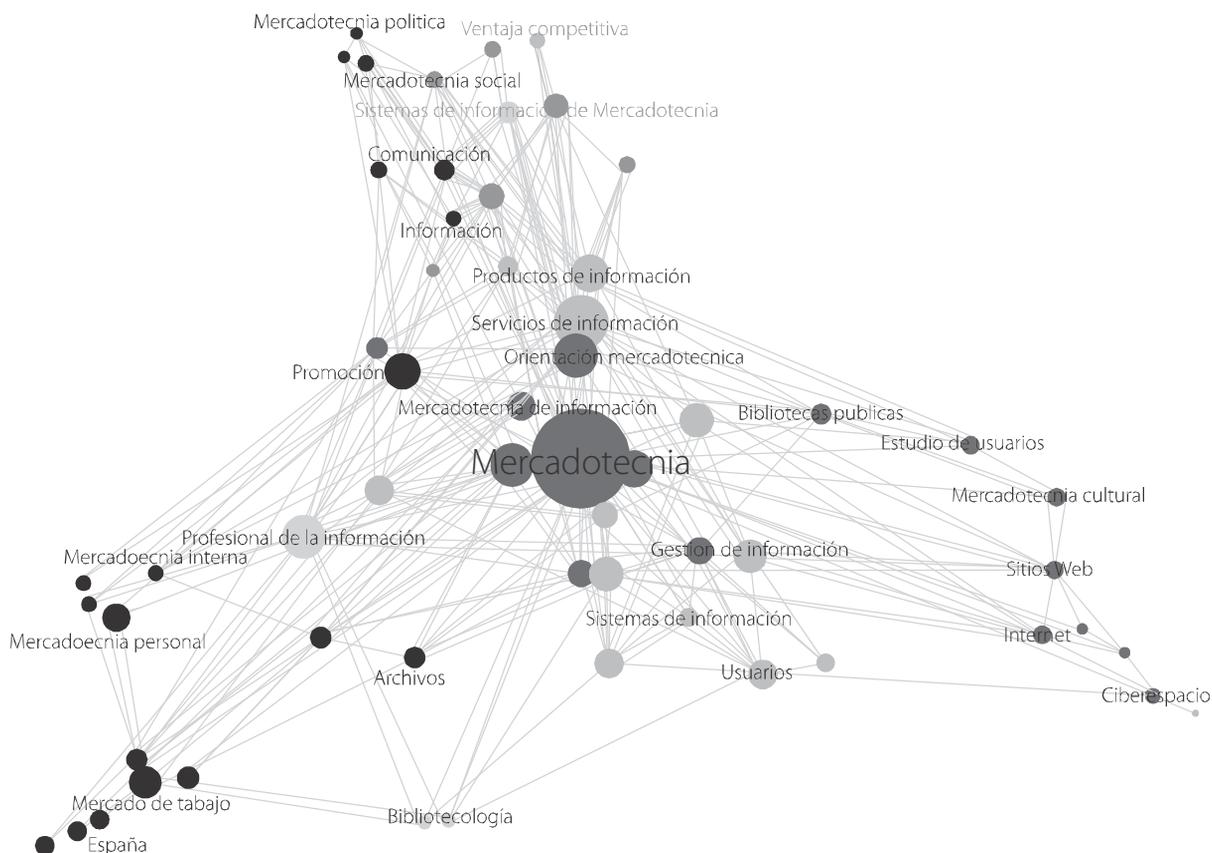


Figura 5. Mapa de coocurrencia de términos.

Fuente: VOSviewer 1.4.0. (2014).

a estos resultados, planteamientos como el de Junqueira y Rodríguez (2009) develan que a lo interno de las relaciones entre el marketing y las CI coexisten indicadores muy comunes como: (1) la dimensión cognitiva y afectiva en la búsqueda de información de los individuos, (2) la manera en la que ellos le dan significado (*sensemaking*) a esa información y al conocimiento del mercado, y (3) el campo de la teoría organizacional; dentro del cual las organizaciones son asumidas como altas consumidoras de información. Estos tres elementos ampliamente recaen bajo la plataforma teórica del comportamiento o conducta informacional que abordan Pettigrew *et al.* (2001).

Recayendo en las técnicas y herramientas mercadotécnicas, la más frecuente es *promoción* (fa: 11), principalmente asociada a *servicios*, *bibliotecas* y *profesional de la información*. Entre las que le siguen están *mezcla de mercadotecnia* (fa: 8), *investigación de mercados* (fa: 7), *mercadotecnia personal* (fa: 5) y *segmentación de*

mercados (fa: 5). Válido es de mencionar la presencia no tan pronunciada de técnicas igualmente contribuyentes como la *mercadotecnia social* (fa: 4), la *política* (fa: 3), la de *relaciones* (fa: 3) y la *interna* (fa: 3). Se distingue además, mediante la figura 2, una asociación precisamente entre técnicas informacionales como *gestión de información* y *gestión del conocimiento* y el tópico *mercadotecnia*; vinculando estas relaciones a otras categorías como *industria de información*, *investigación de mercados*, *unidades de información*, *profesional de la información* y *mercadotecnia de información*.

Conclusión

En sentido general, la mercadotecnia como área del conocimiento dentro de las disciplinas informativas, es una temática que no ha estado en decrecimiento por parte de las publicaciones de la región latinoamericana.

Se considera que uno de los aportes más trascendentales de este artículo radica precisamente en la caracterización de un tópico, a partir de artículos contenidos en un grupo de revistas (51.4%) que no están indizadas por sistemas de información científica de la corriente principal como *Web of Science*, *Scopus*, *Social Science Citation Index*, *Journal Citation Report*, etc. De igual manera, desde Latindex, se han detectado notables índices de colaboración respecto a la temática aquí examinada.

Los resultados han mostrado que Brasil es quien actualmente figura como el país más productivo en cuanto a idioma, autores y revistas. Por otro lado, el análisis de los términos coocurrentes sirvió para identificar cuatro líneas básicas de investigación que se han estado tratando: la mercadotecnia del profesional de la información en el mercado de trabajo, la mercadotecnia en bibliotecas, la mercadotecnia de productos y servicios de información, y la gestión de información de mercadotecnia. Ha sido muy relevante, no solo la exploración de técnicas de mercadeo en el contexto de trabajo de las organizaciones de información, sino también la adopción de técnicas informacionales en la ejecución de tareas mercadotécnicas. También es considerable el alto interés hacia la figura de los profesionales de la información, específicamente porque,

mediante la perspectiva del *marketing* como herramienta gerencial, estos actores logren ganar en relevancia laboral y social; así como en el manejo de nuevas habilidades como factor de cambio en sus disímiles escenarios de actuación.

Latindex, como sistema amplio de información, no posee la infraestructura adecuada para la aplicación de determinados indicadores bibliométricos que podrían resultar útiles en la determinación de algunas tendencias y patrones en la productividad. En añadidura, muchas de las revistas electrónicas identificadas han perdido vigencia (el caso de tres) y otras, al pasar a versión electrónica no brindan la opción de acceder a materiales publicados durante su antigua versión impresa. Con ello se infiere que no se haya trabajado con algunas investigaciones que pudieron haberse desarrollado.

Agradecimientos

Ofrezco mi agradecimiento a la licenciada Anet Ponce Blanco por los apoyos técnicos que amablemente me brindó. De igual manera quiero resaltar las contribuciones de la Dra. C. Nuria Esther Pérez Matos y el Dr. C. Ricardo Arencibia Jorge por las revisiones y sugerencias previas a la publicación de este estudio.

Referencias

Alcain, M.D.; Ascurain, M.L. Análisis de las revistas latinoamericanas de psicología incluidas en el Directorio Latindex. *Papeles del Psicólogo*, n.83, p.34-43, 2002. Disponible en: <<http://www.papelesdelpsicologo.es/vernumero.asp?id=895>>. Acceso en: 30 ago. 2013.

Amaral, S.A. Literatura biblioteconômica brasileira de marketing: produção até 1994. *Revista de Biblioteconomia de Brasília*, v.20, n.1, p.69-98, 1996.

Amaral, S.A. Os 4Ps do composto de marketing na literatura de Ciência da Informação. *Transinformação*, v.12, n.2, p.51-60, 2000.

Cetto, A.M. Ciencia y producción científica en América Latina. El proyecto Latindex. *International Microbiology*, v.1, n.3, p.181-182, 1998. Disponible en: <<http://revistes.iec.cat/index.php/IM/article/viewFile/4c3da6fc9aedf.002/36>>. Acceso en: 30 ago. 2013.

Flores, A.M.; Penkova, S.; Román Román, A. Once años de Latindex: una experiencia al servicio de las publicaciones científicas iberoamericanas. *Simbiosis*, v.6, n.1, p.1-27, 2009. Disponible en: <<http://repositorio.upr.edu:8080/jspui/bitstream/10586/41/3/latindex.pdf>>. Acceso en: 30 ago. 2013.

González-Valiente, C.L.; León Santos, M.S.; Rivera, Z. El egresado de la carrera ciencias de la información y su inserción en la gestión de mercadotecnia. *Revista Cubana de Información en Ciencias de la Salud*, v.25, n.2, 2014. Disponible en: <<http://acimed.sld.cu/index.php/acimed/article/view/583>>. Acceso en: 25 abr. 2014.

Guinchat, C.; Menou, M.; Blanquet, M.F. *Introducción general a las ciencias y técnicas de la información y documentación*. 2.ed. Madrid: Unesco, 1990.

Gupta, D.K. Bibliographical literature on LIS marketing: A review. *Annals of Library and Information Studies*, v.55, n.4, p.308-316, 2006.

Gupta, D.K.; Jain, A.K. Marketing library and information services: A study of periodical literature. *Annals of Library and Information Studies*, v.56, n.4, p.217-226, 2009.

Junqueira, H.L.C.; Rodrigues, R. Busca da informação em marketing: a perspectiva da ciência da informação. *Revista de Administração de Empresas*, v.49, n.2, 2009. Disponível em: <<http://www.redalyc.org/pdf/1551/155113815008.pdf>>. Acceso en: 30 ago. 2013.

Licea de Arenas, J. et al. Una visión bibliométrica de la investigación en bibliotecología y ciencia de la información

de América latina y el Caribe. *Revista Española de Documentación Científica*, v.23, n.1, p.45-53, 2000.

Martínez Rodríguez, A.; Solís Cabrera, F.M. Investigación en el campo de la información en Cuba. Necesidad de su redimensionamiento. *Anales de Documentación*, v.16, n.2, p.1-15, 2013.

Miguel, S.; Moya-Anegón, F.; Herrero-Solana, V. Aproximación metodológica para la identificación del perfil y patrones de colaboración de dominios científicos universitarios. *Revista Española de Documentación Científica*, v.29, n.1, p.36-55, 2006.

Pérez Matos, N. E. *Las disciplinas informativas en Cuba: análisis de su literatura profesional y su relación con los períodos constitucionales de la nación*. 2011. Tesis (Doctoral en Información y Documentación) - Departamento de Biblioteconomía y Documentación, Universidad de Granada, Granada, 2011.

Pettigrew, K.E.; Fidel, R.; Bruce, H. Conceptual frameworks in information behavior. *Annual Review of Information Science and Technology*, v.35, p.43-78, 2001.

Ponjuán D. G. Gestión, gestión de información, gestión del conocimiento... gestión del futuro. *Ciencias de la Información*, v.30, n.3, p.43-52, 1999.

Ríos Gómez, C.; Herrero Solana, V. La producción científica latinoamericana y la ciencia mundial: una revisión bibliográfica

(1989-2003). *Revista Interamericana de Bibliotecología*, v.28, n.1, p.43-61, 2005.

Rodríguez-Gallardo, J.A. La literatura bibliotecológica y de estudios de la información: un análisis crítico de Latindex. *Revista Interamericana de Bibliotecología*, v.32, n.2, p.335-365, 2009.

Schneider, J.W. *Verification of bibliometric methods 'applicability for thesaurus construction*. 2004. Tesis (Doctoral in Information Science) - Department of Information Studies, Royal School of Library and Information Science, Denmark, 2004.

Sieres, E.G. Evaluación de la Revista de Toxicología según los criterios de calidad Latindex. *Revista de Toxicología*, v.23, n.2, p.101-107, 2006. Disponible en: <<http://www.redalyc.org/articulo.oa?id=91923302#>>. Acceso en: 30 ago. 2013.

Verdecia-Gutiérrez, T. Sistema Regional de Información en línea para revistas científicas de América Latina, el Caribe, España y Portugal: un reto para las publicaciones científicas cubanas. *Ciencias de la Información*, v.41, n.1, p.71-73, 2010.

Zins, C. Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and Technology*, v.58, n.4, p.479-493, 2007a.

Zins, C. Knowledge map of Information Science. *Journal of the American Society for Information Science and Technology*, v.58, n.4, p.645-672, 2007b.

Análisis comparativo de herramientas de recuperación y análisis de información de acceso libre desde una concepción docente¹

Comparative analysis of information retrieval and analysis of open access tools from an educational concept

Armando PLASENCIA-SALGUEIRO²

Bárbara de los Milagros BALLAGAS-FLORES³

Resumen

En el Instituto de Cibernética, Matemática y Física de la República de Cuba se imparte el curso "Bases de datos y biblioteca digital" dentro de la Maestría de Cibernética Aplicada. Parte esencial del curso la constituye la creación de bases de datos documentales a partir de la recuperación de información en Internet. Para poder realizar los laboratorios requeridos para un mejor aprendizaje se requiere seleccionar las herramientas de recuperación de información más idóneas, tanto desde el punto de vista docente como desde las facilidades para su adquisición. Se definieron entonces las características para evaluar esas herramientas y la metodología para realizar la selección. Como resultado, de trece herramientas de recuperación y análisis de la información de *software* libre analizadas que pudieron ser descargadas se seleccionaron ocho herramientas, *Lemur Toolkit* con *Indri*, *Sphinx*, *WebSphinx* con *Rapid Miner*, *Solr/Lucene/Hadoop/Mahout*, *Terrier*, *Dragon* lo cual permitió garantizar la calidad del curso impartido y su concatenación con otros cursos de la misma maestría.

Palabras clave: Bases de datos. Minería de textos. Motores de búsqueda. Recuperación de información.

Abstract

In the Institute of Cybernetics, Mathematics, and Physics in the Republic of Cuba the course "Databases and digital Library" is a discipline in the Master's degree program of Applied Cybernetics. An essential part of the course is the creation of documental databases starting from information retrieval from the Internet. To equip the laboratories required for better learning, the most suitable tools for information retrieval are needed, both from an educational point of view as well as the easiness for their acquisition. Therefore, the characteristics to evaluate these tools and the methodology for selecting them were defined. As a result, of the thirteen recovery tools and data analysis from free softwares available to be downloaded, the following eight tools were selected: Lemur Toolkit with Indri, Sphinx, WebSphinx with Rapid Miner, Solr / Lucene / Hadoop / Mahout, Terrier and Dragon, which guaranteed the quality of the course and the connection with other courses in the Master's degree program.

Keywords: Database. Text mining. Searching engines. Information retrieval.

¹ Trabajo presentado en el VII Seminario Internacional sobre Estudios Cuantitativos y Cualitativos de la Ciencia y la Tecnología "Prof. Gilberto Sotolongo Aguilar" en XIII Congreso Internacional de Información - INFO'2014. Habana, Cuba.

² Instituto de Cibernética, Matemática y Física, Departamento de Control Automático. Dirección Postal 10400, La Habana, Cuba. Correspondencia a nombre de/Correspondence to: A.P. SALGUEIRO. E-mail: <armando@icimaf.cu>.

³ Empresa de Telecomunicaciones de Cuba SA., Departamento de Inteligencia Empresarial. La Habana, Cuba.

Recibido el día 10/6/2014 y aceptado para su publicación el 12/9/2014.

Introducción

La Cibernética es la ciencia sobre el control, la obtención, trasmisión y conversión de información en los sistemas cibernéticos. La principal tarea de la Cibernética es la elaboración de la estructura y los métodos de investigación, útiles para el estudio de los sistemas de control independientemente de su naturaleza (Dopico & Plasencia, 2011).

Desde el 2012 se imparte la maestría "Cibernética Aplicada" <www.icimaf.cu/maestria-ca> en el Instituto de Cibernética, Matemática y Física del Ministerio de Ciencia, Tecnología y Medio Ambiente. Esta tiene como objetivo la enseñanza y actualización de los profesionales en las técnicas de descubrimiento de conocimiento basados en datos *Knowledge Data Discovery* (KDD) y su aplicación tanto para la Inteligencia organizacional (en la Mención Minería de Datos de la Maestría), como en la modelación, la identificación, la robótica, el diagnóstico y el diseño de sistemas de control inteligente (en la Mención Control Avanzado).

La concepción del diseño curricular de la maestría se fundamentó en el enfoque sistémico, la enseñanza problémica, la interdisciplinariedad, la multidisciplinariedad, la transdisciplinariedad y la confluencia de las diferentes técnicas para el Control Avanzado y el análisis de datos (minería de datos) (Dopico & Plasencia 2011).

Para respaldar el principio de enseñanza problémica, basado en el criterio de los especialistas al abordar un problema concreto en una determinada organización, la ciencia o la tecnología, se propicia el empleo de herramientas tanto de software libre como propietarias de acceso libre para fines docentes en la realización de laboratorios y ejercicios en los cursos de la maestría con el objetivo también de afianzar los conocimientos teóricos impartidos.

En el curso "Bases de datos y bibliotecas digitales" se imparten entre otros los siguientes temas: recuperación de Información clásica (M1); lenguajes de solicitud de búsquedas (M2); indizado y búsqueda, ficheros invertidos, consultas booleanas, búsquedas secuenciales, reconocimiento de patrones, consultas estructurales, compresión (M3); recuperación de

información multimedia, modelos y lenguajes. Indexado y búsqueda; buscadores y búsqueda en la *Web*.

- *Web Semántica*

- Recuperación de información en redes sociales (M4 y M5)

Se hace necesario entonces, determinar qué herramientas de software se corresponden con los laboratorios y clases prácticas que respondan a los contenidos expuestos con anterioridad.

Sin embargo, ello no es una tarea fácil. La cantidad de herramientas que existen, solo de finales de los 90 a la actualidad es considerable, sus contenidos y alcances varían prácticamente todos los años y para impartir docencia se necesita que éstas estén asequibles en internet a los estudiantes, que estén bien documentadas, que las ayudas sean comprensibles, que existan preferiblemente libros sobre las mismas, que proporcionen ejemplos y estos sean ilustrativos, comprensibles y que su nivel de complejidad sea paulatino.

Por otro lado, las herramientas existentes o son de uso comercial o están muy orientadas a la investigación lo que hace muy confusa la selección de las herramientas adecuadas.

Para darle solución a esta problemática, partimos de la premisa que es posible realizar un análisis cuantitativo y cualitativo de las características de las herramientas para seleccionar las más idóneas para el proceso docente. De ahí que los objetivos que dimanan de esta premisa es el determinar durante la preparación del curso, cuáles son las herramientas que son más adecuadas para la impartición de los laboratorios y establecer una metodología que permita monitorear este proceso de selección.

Trabajos anteriores que aborden esta problemática específicamente desde el punto de vista docente y en una escala de una cantidad suficientemente exhaustiva (más de 10) no hemos encontrado ninguno. Ya desde un punto de vista más general, se destaca el trabajo de (Middleton & Baeza-Yates, 2011) que hace una comparación de 17 motores de búsqueda de acceso libre y la colección de artículos recogidos en la recopilación de trabajos del evento "*Conference on Research and Development in Information Retrieval*" (*Association for*

Computing Machinery/Special Interest Group on Information Retrieval 2012) (Trotman *et al.*, 2012). El inconveniente de estos trabajos es que el primero es del 2007, por lo que estas herramientas han cambiado mucho en siete años y en el segundo caso son trabajos que tratan uno, dos motores de búsqueda.

Recuperación de Información (RI) (Middleton & Baeza-Yates, 2011) es un amplio campo de estudio que abarca la representación, el almacenamiento, la organización y el acceso a los elementos de información. La RI debe de ser capaz de tratar la información de forma tal que se pueda acceder a ella de forma eficiente, enfocada a las necesidades de información del usuario. Una definición más detallada es la de que la RI es la

búsqueda de material de información (generalmente documentos) de naturaleza no estructurada (generalmente textos) que satisface una necesidad de información a partir de una gran colección de datos (generalmente en los servidores de computadoras locales o en Internet).

El proceso de RI se puede desglosar en la interacción de los módulos que se dan en la Figura 1. En esta se destacan tres áreas principales, el indizado, la búsqueda y el ordenamiento:

El análisis consiste en la determinación de nuevos patrones y relaciones entre documentos, no triviales y desconocidos. La interrelación entre la RI y el análisis se puede ver en la Tabla 1.

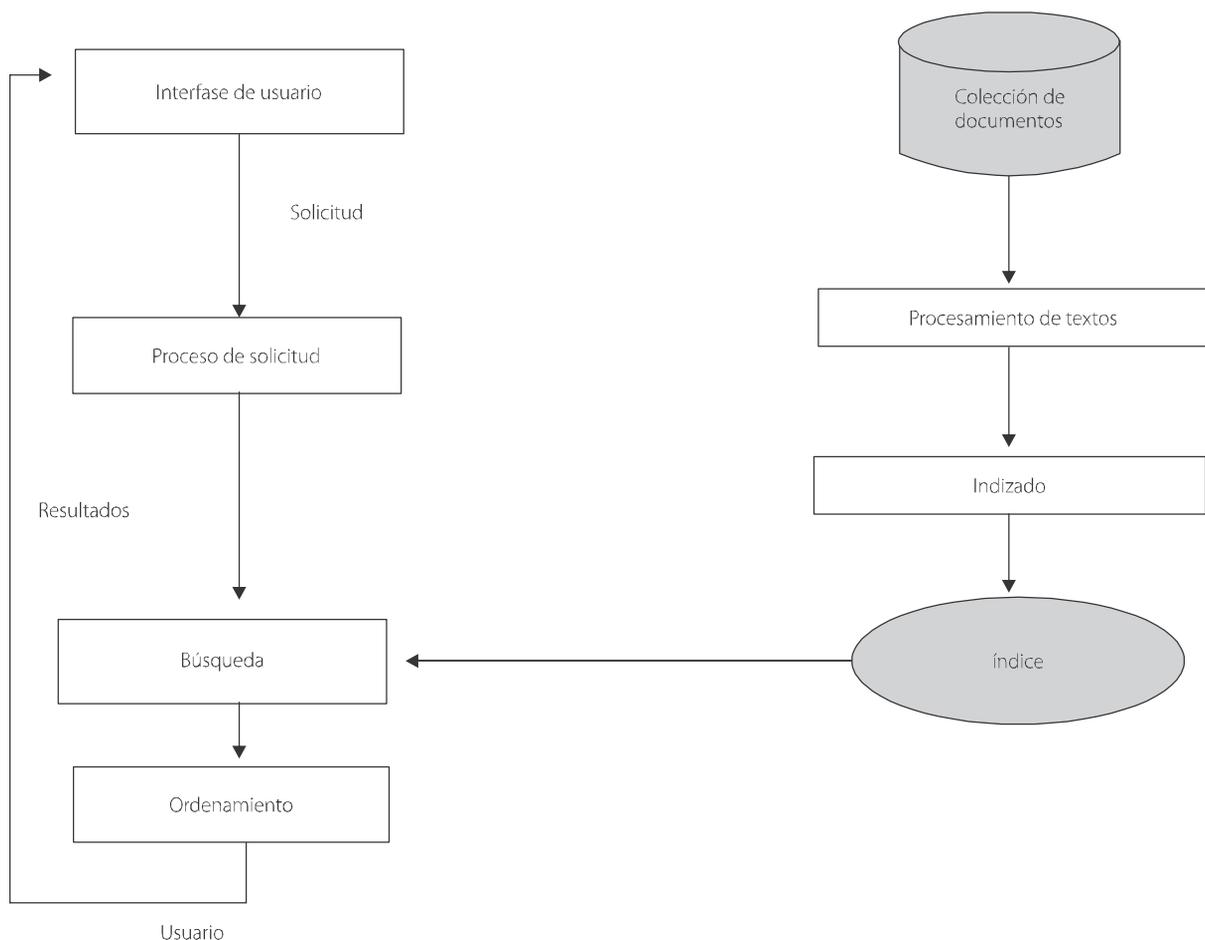


Figura 1. El proceso de recuperación de información.

Fuente: (Middleton & Baeza-Yates, 2011).

Tabla 1. Interrelación entre la recuperación de información y el análisis.

	Fuentes de Datos/Información		
Propósito	Cualquier Dato	Datos textuales	Datos Relativos a la <i>Web</i>
Recuperación de datos o documentos eficiente y efectivamente	Recuperación de datos/ Bases de datos	Recuperación de Información	Recuperación de la <i>Web</i>
Hallazgo de nuevos patrones o conocimiento desconocido con anterioridad para el sistema	Minería de datos	Minería de textos	Minería <i>Web</i>

Fuente: <www.knowledgetechnologies.net/proceedings/presentations/treloar/nathantreloar.ppt>, 2008.

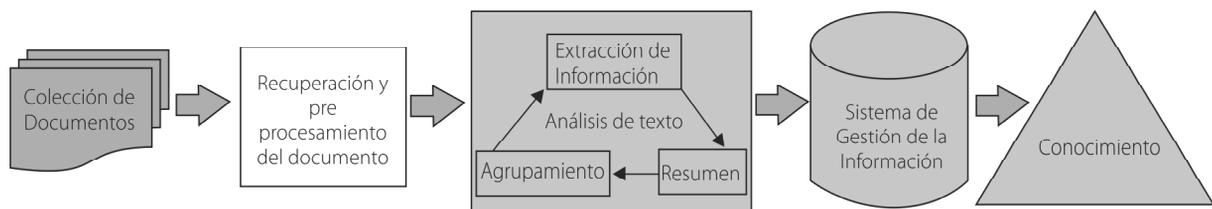


Figura 2. Minería de textos.

Fuente: (Fan *et al.*, 2005).

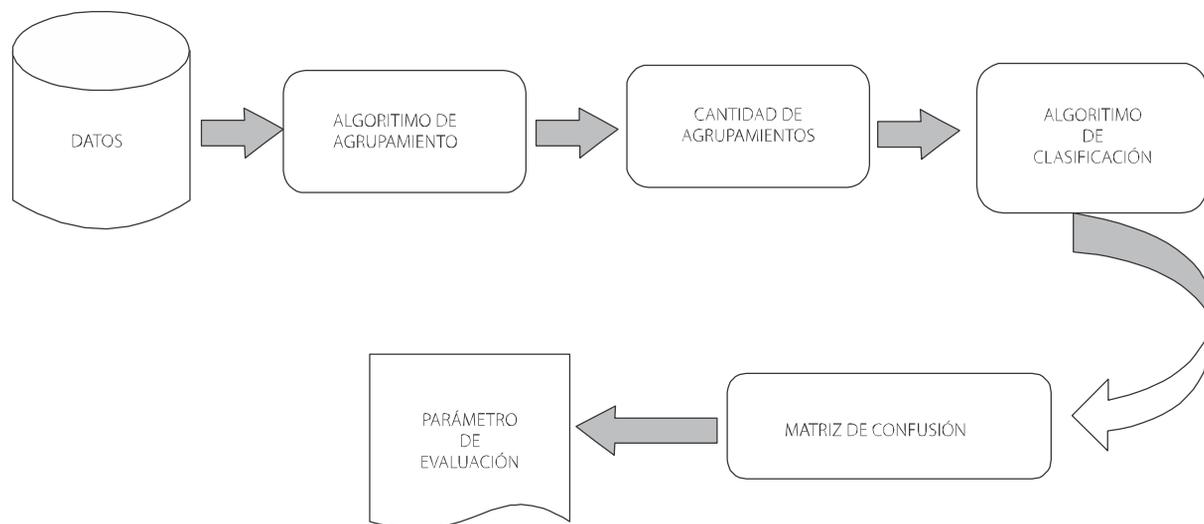


Figura 3. Técnicas de agrupamiento-clasificación.

Fuente: (Pathrey *et al.*, 2013).

Se entiende por minería de textos al proceso de extracción automática de información no trivial, útil, previamente desconocida y finalmente comprensible de

los depósitos de documentos textuales. La interrelación de la recuperación de información con el análisis para obtener nuevo conocimiento se da en la Figura 2.

Dentro de la minería de textos, los tipos de algoritmos que más se emplean son los de agrupamiento y de clasificación, en un proceso relacionado en donde el algoritmo de agrupamiento generalmente es un proceso no supervisado y la clasificación es un proceso supervisado como se muestra en la Figura 3.

Métodos

Para realizar la investigación se hizo un análisis de 13 herramientas de recuperación y análisis. Los criterios que se tomaron para la selección fueron el de que fuesen herramientas de Recuperación de información y/o análisis de información mediante algoritmos de minería de textos (aprendizaje automático) de *software* libre, con código abierto, que independientemente de ser libres no tuviesen restricciones en su descarga, lo cual se verificó mediante la descarga física de los *software*.

Las características de las herramientas se sometieron a un análisis cuantitativo y cualitativo. Para el análisis cuantitativo se utilizaron los parámetros de las comparaciones entre herramientas de *software* que establece el sitio *Web* (Findthebest, 2014). Estos resultados tienen la ventaja de que son contrastables por los usuarios del sitio, dinámicos y de fácil obtención. Su deficiencia radica en que no son transparentes para el usuario en relación con la procedencia inicial de los mismos como si lo pueden ser, por ejemplo, los resultados obtenidos a partir de la ejecución sobre colecciones de datos. Pero esa metodología está fuera del alcance de los recursos y de la propia investigación. Otra deficiencia es que no se incluyen todas las herramientas seleccionadas.

Entre las características que entrega el sitio tenemos: características generales de la herramienta, uso previsto, multiplataforma, multilingüe, identificación conceptual, categorización por relevancia, especificaciones de *software* (licencia, sistema operativo, código programado en, enlace de descarga); y se añaden por los autores los siguientes indicadores: Herramientas de Análisis, Redes Sociales, Manejo información multimedia, *Web* Semántica.

Cuyos datos se obtienen de la información técnica de los sitios de hospedaje de las herramientas.

Se incorporan también elementos cualitativos de análisis como resultado de la experiencia de los autores en la impartición del curso durante tres sesiones y la interacción con los alumnos en los laboratorios. Estos son: Especificaciones docentes: facilidad en la ejecución de los algoritmos de RI, facilidad de ejecución de los algoritmos de análisis, visualización de los resultados, comunidad de discusión, facilidad de la ayuda, existencia de ejemplos, Libro/Manual/Idioma, empleo en los cursos.

Resultados y Discusión

Relación de las herramientas de recuperación y análisis seleccionadas

Toolkit <<http://www.lemurproject.org/>>: Lemur es una herramienta destinada para facilitar la investigación en la modelación del lenguaje y en la recuperación de información. Para ello incluye tecnologías como recuperación ad hoc y distribuida, consultas estructuradas, RI multilingüe, indización general, filtrado y categorización. Proporciona aplicaciones de ejemplo diseñándose la herramienta para permitir la fácil programación de sus propias actualizaciones y aplicaciones.

Características (Findthebest, 2014): sofisticado lenguaje estructurado de consulta; recuperación de documentos estructurados y en XML; indizado de las páginas *Web* con capacidad de búsqueda de sitios "out-of-the-box"; interfases interactivas para Windows, Linux y *Web*; múltiples métodos de indizado para colecciones pequeñas, medianas y grandes (*terabytes*). Indexado incremental; realimentación relevante y pseudo relevante, recuperación multilingüe. API en C++. Java y C#; recuperación de información distribuida y aplicaciones de agrupamiento de documentos.

Sphinx <<http://sphinxsearch.com/>>. *Sphinx* es un motor de búsqueda de acceso abierto orientado en su diseño a la indización del contenido de las bases de datos. De forma nativa incorpora como gestores de bases de datos a *MySQL* y a *PostgreSQL*. Otras fuentes de datos pueden ser indexadas mediante conexión en un formato de edición en XML.

Características (Findthebest, 2014): alta velocidad de indizado y búsqueda; incluye un ordenamiento por proximidad de frases, proporcionando una buena relevancia; también incorpora cualquier número de campos de documentos y tratamiento de las palabras vacías.

The Dragon Toolkit <<http://dragon.ischool.drexel.edu>>. Descripción (Zhou *et al.*, 2007): es un paquete de desarrollo escrito en Java para uso académico en RI y minería de textos *Text Mining* (TM) (incluye clasificación, agrupamiento, resumen y modelación). Está destinado a los investigadores que trabajan con tareas de RI y TM a gran escala y prefieren la programación en Java. Es diferente de Lucene y de Lemur, proporcionando un soporte incorporado para RI y TM basado en semántica. No incluye RI distribuida ni es multilingüe lo cual es propio de Lucene y Lemur. Incluye la obtención de ontologías.

Otra característica importante es su escalabilidad y la inclusión de herramientas de programación en lenguaje natural *Natural Language Programming* (NLP).

Text-Garden - Text-Mining Software Tools <<http://ailab.ijs.si/dunja/textgarden/>>. Descripción (Grobelnik, 2007): la herramienta de TM *Text-Garden* posibilita la fácil manipulación de los documentos textuales para su análisis incluyendo la generación automática de modelos y la clasificación, agrupamiento y visualización de documentos; gestiona documentos en la *Web*, rastrea la *Web* y otras muchas funciones. Escrita en código C++ y corre en Windows.

Es un *software* multiplataforma y tiene las siguientes interfases con la misma API: C/C++, Java, .NET, Matlab, Python y R.

Apache Lucene/Solr <<http://lucene.apache.org>>. Descripción: Lucene es una biblioteca de RI escalable, de altas prestaciones (Hatcher *et al.*, 2010). Permite adicionar capacidades de búsqueda a las aplicaciones. Implementada en Java, es un proyecto en la *Apache Software Foundation*.

Las características más importantes de Lucene: un indizado invertido para la recuperación eficiente de los documentos por los términos indizados. La misma tecnología respalda a los datos numéricos al realizarse consultas de rango. Un rico conjunto de componentes de análisis de textos encadenables, tales como

señalizadores y radicalizadores de lenguajes específicos que transforman una cadena textual en una serie de términos. Una sintaxis de consulta con un analizador sintáctico y una variedad de tipos de consulta desde la consulta de un término simple hasta un exótico pareo fuzzy. Un buen algoritmo de posicionamiento basado en principios fundamentados en la Recuperación de Información (IR) para generar primero el candidato más parecido, con significados flexibles para afectar el posicionamiento. Mejoramiento de las características de búsqueda.

Solr <<http://lucene.apache.org/solr/>>. Solr es un servidor de búsqueda empresarial de código abierto. El servidor realiza la comunicación utilizando los estándares *HyperText Transfer Protocol* (http), *eXtensible Markup Language* (XML), y *JavaScript Object Notation* (JSON).

El motor de búsqueda subyacente de Solr es Lucene

Solr tiene características propias que van más allá del propio Lucene (Plasencia *et al.*, 2012). Estas son: ficheros de configuración, en particular por su esquema de índices, los cuales definen los campos y la configuración de sus análisis de textos; un analizador sintáctico de la consulta llamado *dismax* que es más utilizable para el análisis de las consultas del usuario final que el analizador sintáctico de consultas nativo de Lucene; búsqueda geoespacial para el filtrado y el ordenamiento por la distancia; una propiedad de búsqueda distribuida y de replicación del índice para la adaptación de Solr.

Swish-e <<http://swish-e.org/index.html>>. Descripción: Simple *Web* Indexing System for Humans – Enhanced (Swish-e) es un sistema rápido y flexible destinado para la indización de colecciones de páginas *Web* u otros ficheros. Swish-e es ideal para trabajar con colecciones de millones de documentos o menos. Utilizando el analizador sintáctico Gnome™ libxml2 y una colección de filtros, Swish-e puede indizar textos completos, correos electrónicos, *pdf*, *html*, *Microsoft® Word/PowerPoint/Excel* y cualquier fichero que pueda convertirse a texto xml o html. Es utilizado frecuentemente para complementar bases de datos del

tipo MySQL® *Database Management System* (DBMS) con la búsqueda rápida a texto completo.

Características: indizado rápido de un gran número de documentos en diferentes formatos incluyendo textos, html y xml; utiliza "filtros" para indizar otros tipos de ficheros tales como pdf, gzip, o *PostScript*; incluye un rastreador *web* para el indizado de documentos remotos en http. Sigue las reglas de exclusión de los Robots (incluyendo las Meta etiquetas).

Xapian <<http://xapian.org/>>. Descripción: es una biblioteca de un motor de búsqueda de libre acceso, distribuida bajo licencia *General Public License* (GPL).

Esta biblioteca posibilita la creación de una herramienta muy adaptable que permite a los desarrolladores adicionar fácilmente índices avanzados y facilidades de búsqueda a sus propias aplicaciones. Incluye el modelo de RI probabilístico así como un rico conjunto de operadores de consulta booleanos.

La última versión estable es la 1.2.16, distribuida el 2013-12-4

Características: escrita en C++; enlaces que permiten el uso de Php, Perl, Python, C#, Ruby; incluye el modelo de RI probabilístico (Okapi BM25); además de la biblioteca, tiene un número de programas de ejemplo, y una aplicación desarrollada para el indizado y la búsqueda (Omega).

Web SPHINX <<http://www.cs.cmu.edu/~rcm/websphinx/>>. Descripción: *Website-Specific Processors for html Information Extraction (Web SPHINX)* es una biblioteca de clases de Java y un entorno interactivo para los rastreadores *Web*. Un rastreador *Web* (también denominado robot o araña) es un programa que navega y procesa las páginas *Web* de forma automática (Miller & Bharat, 1998).

Web SPHINX está compuesto de dos partes el *Crawler Workbench* y la biblioteca de clases de *Web SPHINX*.

El *Crawler Workbenches* es una interfase gráfica al usuario que permite controlar y adaptar el buscador *web*, un número determinado de operaciones comunes de acceso están incorporadas al rastreador posibilitando al usuario especificar y correr rastreos simples sin necesidad de programas.

La biblioteca de clases de *Web SPHINX* proporciona los elementos para escribir rastreadores de la *Web* en Java.

RapidMiner <www.rapid-i.com>: Descripción: *RapidMiner/Yale*: Líder en las herramientas de minería de datos debido a la combinación de su tecnología y del umbral de funcionalidades disponibles. Cubre una amplia gama de algoritmos de Minería de Datos (MD) incluyendo los algoritmos nativos de *Waikato Environment for Knowledge Learning* (WEKA). Además de ser una herramienta flexible para aprender y explorar la minería de datos, la interfaz gráfica de usuario tiene como objetivo simplificar el uso para las tareas complejas de esta área (Cardoso *et al.*, 2011).

Características de *RapidMiner*: *RapidMiner* es un sistema prototipado para el descubrimiento del conocimiento y Data Mining; usa el lenguaje de scripting xml para describir los operadores y su configuración; posee una interfaz gráfica, línea comando, y API de Java para usar *RapidMiner* desde los propios programas; una gran cantidad de extensiones (plugins); las aplicaciones incluyen: *Text Mining*, *Multimedia Mining*, entre otras.

El *TextInputPlugin* de *Rapid Miner* proporciona un rastreador *Web* simple sobre la base de *WebSPHINX Open Sourcecrawler* escrito en Java.

Indri <<http://www.lemurproject.org/>>. Descripción: el lenguaje de consulta *Indri*, basado en el lenguaje de consulta *Inquery* fue diseñado para ser un lenguaje robusto. Este puede trabajar tanto con una consulta simple como con una consulta muy compleja. Tal lenguaje de consulta sitúa a *Indri* en un lugar diferente del resto de los motores de búsqueda. Éste permite la comparación con frases complejas, sinónimos, expresiones ponderadas, filtrado booleano, campos numéricos (y fechas) y el uso extensivo de documentos estructurados (campos), entre otros.

Indri es un motor de búsqueda de textos desarrollado en la *University of Massachusetts*. Éste es parte del proyecto *Lemur*.

Por último, si *Lucene/Solr* es atractivo porque es un paquete con casi todas las características necesarias. Para realizar aplicaciones basadas en la *Web*, *Indri* es atractivo porque ofrece mejores resultados de búsqueda y como lenguaje de consulta altamente expresivo, lo que

permite una buena granulación del control de la búsqueda (Turtle *et al.*, 2012).

TERRIER <<http://terrier.org>>. Descripción (Middleton & Baeza-Yates, 2011): TERabyte RetriEveR (TERRIER) es una plataforma modular que permite un rápido desarrollo de los motores de búsqueda para la Web, la intranet o las PC desarrollado por la Universidad de Glasgow en Escocia. Éste tiene la posibilidad de indizar, consultar y evaluar las colecciones estándar de *Text REtrieval Conference* (TREC).

Características: aplicable a grandes colecciones de documentos; disponible como una aplicación de la PC, una interfase *Web Java Server Page* (JSP) y una *Application Programming Interface* (API); expansor de acrónimos; posibilita el indizado de textos, html, PDF, Microsoft Word, Excel, PowerPoint, y colecciones TREC; indizado de la información de campo (ejemplo, frecuencia de términos en el campo title); recuperación en la PC, líneas de comando y la Web basada en interfaces de consulta; varios modelos de ponderación de los documentos, Okapi BM25, modelación del lenguaje y tf-idf; facilidades de expansión de la consulta mediante realimentación seudo relevante; lenguaje de consulta avanzado basado en operadores booleanos, +/-, búsqueda por frases y por proximidad y búsqueda por campos.

Hadoop/Mahout/Solr. Descripción (Plasencia *et al.*, 2012): *Hadoop* <<http://hadoop.apache.org/>>. Las bibliotecas de programas de Apache Hadoop constituyen un marco de trabajo que permite el procesamiento distribuido de grandes conjuntos de datos a través de clusters de computadoras utilizando un modelo simple de programación. Formalmente, Hadoop es un marco de trabajo de código abierto para la escritura y la corrida de aplicaciones distribuidas que procesan grandes cantidades de datos. Soporta el modelo MapReduce.

Apache Mahout <<http://mahout.apache.org/>>: Mahout es una biblioteca de aprendizaje automático de Apache. Los algoritmos que éste implementa caen bajo la esfera del aprendizaje automático o inteligencia colectiva. Ello significa en particular para Mahout el disponer de motores de recomendación (filtrado colaborativo), agrupamiento y clasificación. Éste es también escalable. Mahout tiene la finalidad de ser una herramienta de aprendizaje automático a seleccionar

cuando la colección de datos a ser procesados es muy grande para una sola máquina.

Weka4WS/Globus Toolkit 4 (GT4) <<http://grid.deis.unical.it/papers/pdf/PKDD2005.pdf>>. Descripción: MapReduce no es el único método donde el aprendizaje automático distribuido tiene lugar, hay adiciones a Weka, tal como Weka4WS que están disponibles y realizan en esencia las mismas prestaciones. Weka4WS permite realizar el procesamiento distribuido de tres tareas comunes en Weka: el etiquetado, donde las etiquetas de las clases son asignadas a las instancias basadas en un modelo predictivo, prueba, donde es calculado el algoritmo de precisión de la predicción mediante el etiquetado de un conocido conjunto de instancias y la validación cruzada, un método de prueba que divide un conjunto de datos de entrada en n particiones. Estos tres métodos son ideales para una distribución simple de un conjunto de datos dividido utilizado a través de múltiples máquinas que no cambian el resultado final (Jenkin, 2009).

Weka4WS ha sido desarrollado utilizando la biblioteca de Java proporcionada por *GlobusToolkit 4* (GT4), el cual tiene una arquitectura *Open Grid Service Architecture* (OGSA) (Bernal, 2008).

A continuación se muestran las dos tablas obtenidas de la comparación de las 13 herramientas de recuperación y análisis de información (Tabla 2).

Ahora podemos determinar cuáles son las herramientas más adecuadas para la realización de los laboratorios.

Para trabajar el Módulo 1, que incluye la introducción a las bases de datos documentales y las bibliotecas digitales, sus conceptos básicos, el proceso de recuperación de información, la modelación, la recuperación: Adhoc y el filtrado, la caracterización formal de los modelos de RI, la recuperación de información clásica, la evaluación de la recuperación y las colecciones de referencia, se seleccionó la herramienta Lemur Toolkit con Indri por la documentación en línea, la comunidad de discusión, la cantidad de ejemplos, la ayuda del programa y la utilización en otros cursos de diferentes universidades.

Para el Módulo 2, que comprende los lenguajes de solicitud de búsquedas, las solicitudes basadas en

Tabla 2. Comparación de los programas de recuperación de información.

1 de 1

Programas de Recuperación de Información (A)							
Características	Programa						
	Lemur Toolkit	Sphinx	Dragon	Text Garden	Lucene	Swish-e	
Uso previsto	Motor de búsqueda	Motor de búsqueda	Biblioteca	Biblioteca	Biblioteca	Buscador	
Multiplataforma	Si	Si	N/A	Si	Si	N/A	
Multilingue	Si	No	No	No	N/A	N/A	
Identificación conceptual	Si	N/A	N/A	N/A	N/A	N/A	
Categorización por relevancia	Si	Si	Si	Si	Si	Si	
Herramientas de Análisis	No	No	Si	Si	No	No	
Redes Sociales	No	No	No	No	No	No	
Manejo información multimedia	No	No	No	No	No	No	
Web Semántica	No	No	Si	No	No	No	
Especificaciones Software							
Licencia	Código libre/ Gratis	Código libre	Código libre	Código libre	Código libre	Código libre	
Sistema operativo	Linux/Windows	Linux/Windows	Linux/Windows	Windows	Linux/Windows	Linux/Windows	
Código programado en:	C/C++	C/C++	Java	C/C++	Java	C/C++	
Enlace de descarga	< http://sourceforge.net/projects/lemur/ >	< http://www.sphinxsearch.com/downloads.html >	< http://dragon.ischool.drexel.edu >	< http://ailab.ijs.si/dunja/textgarden/Txt2Cpd.zip >	< http://lucene.apache.org/ >	< http://swish-e.org/download/index.html >	
Versión	4.12	0.9.9	1.3	N/A	4.6	2.4.7	
Especificaciones Docentes							
Facilidad en la ejecución de los algoritmos de RI	Baja	Baja	Baja	Baja	Media	Baja	
Facilidad de ejecución de los algoritmos de análisis	Baja	Baja	Baja	Baja	N/a	Baja	
Visualización de los resultados	No	No	No	Si	N/a	No	
Comunidad de discusión	Si	N/A	N/A	N/A	Si	N/A	
Facilidad de la ayuda	Media	Alta	Media	Media	Si	Media	
Existencia de ejemplos	Si	Si	Si	Si	Si	Si	
Libro/Manual/Idioma	No	Si (pdf)/inglés	Si (pdf)/inglés	No	Si (pdf)/inglés	No	
Empleo en Cursos	Si	N/A	N/A	N/A	Si	N/A	
Programas de Recuperación de Información (B)							
Características	Programa						
	Xapian	WebSphinx	RapidMiner	Indri	Terrier	Hadoop/Mahout	Weka4WS
Uso previsto	Biblioteca	Biblioteca	Programa	Buscador	Motor de búsqueda	Biblioteca	Programa
Multiplataforma	N/A	N/A	Si	Si	N/A	Si	Si
Multilingue	N/A	N/A	No	No	Si	N/A	N/A

Tabla 2. Comparación de los programas de recuperación de información.

	Programas de recuperación de información (B)						
	Xapian	WebSphinx	RapidMiner	Indri	Terrier	Hadoop/Mahout	Weka4WS
<i>Programa</i>							
<i>Características</i>							
Identificación conceptual	N/A	N/A	No	Si	Si	N/A	N/A
Categorización por relevancia	Si	No	No	Si	Si	N/A	N/A
Herramientas de análisis	No	No	Si	No	No	Si	Si
Redes Sociales	No	No	Si	No	Si	N/A	N/A
Manejo información multimedia	No	No	Si	No	Si (con Smart)	N/A	N/A
Web semántica	No	No	Si	No	N/A	N/A	N/A
<i>Especificaciones software</i>							
Licencia	Código libre	Código libre	Código libre	Código libre	Código libre	Código libre	Código libre
Sistema operativo	Windows	Windows	Linux/ Windows	Linux/ Windows	Linux/ Windows	Linux/ Windows	Linux
Código programado en:	C/C++	Java	Java	C/C++	Java	Java	Java
Enlace de descarga	< http://xapian.org >	< http://www-2.cs.cmu.edu/~rcm/websphinx/ >	< http://rapidminer.com/ >	< http://sourceforge.net/projects/lemur/files/latest/download?source=files >	< http://terrier.org/download/agree.shtml?terrier-3.5.tar.gz >	< ">https://cwiki.apache.org/confluence/display/MAHOUT/Downloads.>	< http://grid.deis.unical.it/papers/pdf/PKDD2005.pdf >
Versión	1.2.16	N/A	6.0	5.6	3.5	0.5	N/A
<i>Especificaciones Docentes</i>							
Facilidad en la ejecución de los algoritmos de RI	Baja	Media	Alta	Baja	Media	Media	Baja
Facilidad de ejecución de los algoritmos de análisis	Baja	N/A	Alta	Baja	N/A	Media	Media
Visualización de los resultados	No	Si	Si	No	Si	Si	Si
Comunidad de discusión	Si	No	Si	Si	Si	Si	N/A
Facilidad de la ayuda	Media	Baja	Alta	Media	Media	Alta	Baja
Existencia de ejemplos	Si	No	Si	Si	Si	Si	N/A
Libro/Manual/Idioma	No	Si	Si	No	No	Si (pdf)/inglés	N/A
Empleo en Cursos	N/A	Si	Si	Si	N/A	Si	N/A

Fuente: Elaboración propia (2014).

Nota: N/A: No disponible; RI: Recuperación de Información.

palabras claves, el reconocimiento de patrones, los protocolos de solicitud, los lenguajes textuales y propiedades, metadatos, textos, lenguajes de marcación,

multimedia, operaciones con textos, preprocesamiento de documentos, agrupamiento, compresión de datos, se selecciona Sphinx por su semejanza con Lemur pero por

incorporar además la facilidad de trabajar con PostgreSQL como gestor de bases de datos no estructuradas, lo que le da continuidad al curso de “Bases y almacenes de Datos” que se imparte con anterioridad en la misma maestría y por disponer de una buena documentación.

Para el Módulo 3, en donde se imparte el indizado y la búsqueda, los ficheros invertidos, las solicitudes booleanas, las búsquedas secuenciales, las solicitudes estructurales, la recuperación de información paralela y distribuida, las interfases al usuario y su visualización, la interacción hombre-computadora, el proceso de acceso de información, puntos de inicio, las especificaciones de las solicitudes, contexto, utilización de los juicios de relevancia y las interfases para el proceso de búsqueda, se selecciona *WebSphinx* con *Rapid Miner*, que une, en dos herramientas gráficas las utilidades de rastreo en la Web con las de análisis de los datos con minería de texto en formato XML y da continuidad al curso de “Recuperación de Información y Minería Web” que se imparte en la misma maestría. Para mostrar el trabajo con el procesamiento distribuido se selecciona el conjunto de herramientas “Solr/Lucene/Hadoop/Mahout” por su excelente literatura, comunidad de discusión, facilidad de descarga y el poder mostrar los ejemplos tanto en un cluster, una grid o una máquina personal.

Por último, para el Módulo 4, que incluye la recuperación de información multimedia, los buscadores y la búsqueda en la Web. Se propone el empleo del motor de búsqueda Terrier para el trabajo con redes sociales y datos multimedia (en conjunto con Smart) y la biblioteca Dragon para el trabajo con ontologías y la Web semántica.

El método expuesto de seguimiento y actualización de la Tabla 2, proporciona una metodología a seguir para el desarrollo y la evolución de las herramientas de recuperación de información, lo cual es

un proceso que se debe de realizar en la preparación de cada curso si se quieren impartir contenidos actualizados a los estudiantes y elevar así la calidad de los mismos.

Conclusión

Como resultado del análisis técnico y cualitativo de acuerdo con la experiencia de los profesores del curso “Bases de datos documentales y biblioteca digital” que se imparte en la mención de Minería de datos de la maestría “Cibernética Aplicada” en el Instituto de Cibernética, Matemática y Física, en La Habana, Cuba, de trece herramientas de recuperación y análisis de la información de software libre analizadas que pudieron ser descargadas libremente de Internet se seleccionaron ocho herramientas, LemurToolkit con Indri, Sphinx, WebSphinx con Rapid Miner, Solr/Lucene/Hadoop/Mahout, Terrier y Dragon lo que permite realizar un análisis comparativo entre las herramientas e incorporar la enseñanza problemática dentro de las actividades del curso para un mejor proceso de aprendizaje.

Se establece una metodología para el seguimiento y selección de nuevas herramientas ante el comienzo de un nuevo curso.

La comprensión del funcionamiento y el dominio del uso de estas herramientas es de gran ayuda a los profesionales de la información que en ocasiones realizan el trabajo de recuperación de la información de forma manual o semiautomática, o que simplemente no pueden realizar la recuperación de la información en grandes volúmenes de datos, como pueden ser los que se originan de fuentes como las redes sociales.

Los datos expuestos en la tabla son susceptibles de revisión así como las características a incorporar a la misma, para que reflejen mejor las cualidades de las herramientas mostradas. Ello deberá realizarse de forma continuada en los próximos cursos.

Referencias

Bernal, J. *Data mining and cross-validation over distributed: Grid enabled networks: Current state of the art*. Florida: Atlantic University Spring, 2008. Available from: <<http://latina.mericangrid.org/elgg/juan.bernal/files/2/13/Project+-+DataMining+-+CrossValidation+in+Grid+Enabled+Networks.ppt>>. Cited: Dec. 20, 2013.

Cardoso, Y. *et al. Herramientas de minería de datos*. 2011. Disponible en: <<http://www.monografias.com/trabajos92/herramientas-mineria-datos/herramientas-mineria-datos.shtml>>. Acceso en: 13 enero. 2014.

Dopico, I.; Plasencia, A. Diplomado control avanzado: pertinencia y concepción curricular. In: Convención y Feria

Internacional Informática, 24., 2011. La Habana. *Resumen...* La Habana: CLAD, 2011. p.5-6.

Fan, W. *et al.* Tapping into the power of text mining. *Communications of the ACM*, v.49, n.9, p.77-82, 2005. Available from: <http://filebox.vt.edu/users/wfan/paper/text_mining_final_preprint.pdf>. Cited: Jan. 13, 2014.

FindTheBest. *Compare full text search software*. 2014. Available from: <<http://full-text-search.findthebest.com/>>. Cited: Jan. 13, 2014.

Grobelnik, M. *Text-Garden software suite quick overview*. Ljubljana, Slovenia: Jozef Stefan Institute. 2007. Available from: <http://www.powershow.com/view1/f5de0-ZDc1Z/TextGarden_Software_Suite_Quick_Overview_powerpoint_ppt_presentation>. Cited: Sep. 17, 2009.

Hatcher, E.; Gospodnetic, O.; McCandless, M. Lucene in action. 2nd ed. 2010. *E-book*. Stamford: Manning Publications. Available from: <dl.e-book-free.com/2013/07/lucene_in_action_2nd_edition.pdf>. Cited: Jan. 13, 2014.

Jenkin, N. *Distributed machine learning with Hadoop*. 2009. Disponible en: <http://wenku.it168.com/d_000575816.shtml> Acceso en: 13 enero 2014.

Middleton, C.; Baeza-Yates, R. *A comparison of open source search engines*. Barcelona: Universitat Pompeu Fabra, 2011. Available from: <<http://wrg.upf.edu/WRG/dctos/Middleton-Baeza.pdf>>. Cited: Jan. 13, 2014.

Miller, R.; Bharat, K. SPHINX: A framework for creating personal, site-specific web crawlers. In: International World Wide Web Conference, 7., 1998. Brisbane, Australia. *Proceedings...* Brisbane, Australia: Computer Network and ISDN Systems, v.30, p.119-130, 1998.

Pathrey, R. *et al.* Discovering knowledge patterns from Integration of clustering and classification techniques. *International Journal of Advanced Research in Computer Science and Software Engineering*, v.3, n.4, p.338-343, 2013.

Plasencia, A. *et al.* *Concepción de un buscador web soportado en tecnología grid e interrelación de herramientas Apache*. La Habana: Instituto de Cibernética Matemática y Física, 2012.

Trotman, A. *et al.* Towards an efficient and effective search engine. In: International ACM SIGIR Conference on Research on Development in Information Retrieval, 35., 2012, Portland. *Proceedings...* Portland: University of Otago, 2012, p.40-47.

Turtle, H.; Hegde, Y.; Rowe S. Yet another comparison of Lucene and Indri performance. In: International ACM SIGIR Conference on Research on Development in Information Retrieval, 16., 2012, Portland. *Proceedings...* Portland: University of Otago, 2012, p.64-67.

Zhou, X.; Zhang, X.; Hu, X. *The dragon toolkit developer guide*. 2007. Philadelphia: Drexel University. Available from: <<http://dragon.ischool.drexel.edu/tutorial.pdf>>. Cited: Jan. 13, 2014.

Análisis de contenido del dominio tecnológico *vegetable oil combustion*¹

Content analysis of the technological domain vegetable oil combustion

Noé PÉREZ-ARREORTÚA²

Maidelyn DÍAZ-PÉREZ³

Raudel GIRÁLDEZ-REYES³

Humberto Andrés CARRILLO-CALVET⁴

Resumen

La patente es una medida de productividad científico-tecnológica muy utilizada como indicador tecnológico. Sus análisis está enfocado principalmente a recuentos por año, países, titulares e inventores, pero escasean los estudios de patentes considerando la clasificación técnica utilizada y la información contenida en el documento de invención. La presente investigación tiene como objetivo proponer el uso de la clasificación internacional de patentes y el análisis de contenido de determinados campos del documento de patente como unidad de análisis y medida para realizar un estudio patentométrico en el dominio tecnológico *vegetable oil combustion*. Se utiliza un conjunto de indicadores simples y relacionales, y el *software proINTEC* para el análisis y representación de los datos. La interpretación de los resultados confirma que la información contenida en las patentes de este dominio es pertinente a los intereses del proyecto que ejecuta el caso de estudio.

Palabras clave: Aceite vegetal. Análisis de contenido. Clasificación internacional de patentes. Patentes. Combustión.

Abstract

The patent document is a measurement unit for scientific and technological productivity commonly used as a technological indicator. Patent studies mainly focus on frequency counts by year, country, assignee and inventor. There are scarce studies on the technical classifications and contents of the invention. The aim of this paper is to propose the use of the International Patent Classification and content analysis of some fields of the patent document as a unit of measurement and analysis to conduct a study on patent metrics in the technological domain vegetable oil combustion. A set of simple and relational indicators were used and the software proINTEC analyzed the data. The interpretation of the results proves that the information contained in patent documents from this domain is relevant for the goals of the project in the present case study.

Keywords: Vegetable oil. Content analysis. International patent classification. Patents. Combustion.

¹ Trabajo presentado en el VII Seminario Internacional sobre Estudios Cuantitativos y Cualitativos de la Ciencia y la Tecnología "Prof. Gilberto Sotolongo Aguilar" en XIII Congreso Internacional de Información - INFO' 2014. Habana, Cuba.

² Universidad La Salle Oaxaca, Escuela de Ingeniería. Oaxaca, México.

³ Universidad de Pinar del Río, Grupo Gestión de Información, Conocimiento y Tecnologías, Dirección de Información Científico Técnica. Calle Martí Esquina #300, entre 27 de Noviembre y González Alcorta, 20100, Pinar del Río, Cuba. Correspondencia a nombre de/Correspondence to: M. DÍAZ-PÉREZ. E-mail: <maidelyn@ict.upr.edu.cu>.

⁴ Universidad Nacional Autónoma de México, Facultad de Ciencias, Departamento de Matemática, Laboratorio de Dinámica no Lineal. Ciudad de México, México.

Recibido el día 27/5/2014 y aceptado para su publicación el 9/9/2014.

Introducción

La patente es una medida de productividad científico-tecnológica, es un documento que comunica el avance científico al igual que los artículos. Las patentes, incluso indican que la tecnología resultante puede ser aplicada en los procesos de producción, dato que influye y facilita un mejor desempeño en las transferencias de tecnologías y licencias tecnológicas. Las patentes son un indicador tecnológico de significativa relevancia en los estudios de innovación tecnológica (Basberg, 1982, 1983; Pavitt, 1985, 1988; Grupp *et al.*, 1991; Narin, 1995), y una de las fuentes de datos más confiables para realizar estudios bioinformétricos en las últimas décadas (Rip & Courtial, 1984; Leydesdorff & Heimeriks, 2001; Díaz-Pérez & Moya-Anegón, 2008). Por ello hoy, dentro de los análisis bibliométricos se incluyen los estudios con indicadores de patentes (Narin & Olivastro, 1988; Narin, 1994; Spinak, 1998; Klitkou & Gulbrandsen, 2010; Pao-Long *et al.*, 2010).

A pesar de ser tan importantes los estudios de patentes, la mayoría de sus análisis trabajan recuentos por años, países, titulares e inventores, escaseando los estudios patentométricos que realicen análisis con el contenido del documento de invención.

Dentro de los diferentes tipos de mapas que se pueden elaborar mediante análisis métricos de información, uno de los más estudiados es el mapa de coocurrencias de palabras. Se elabora cuando se desea reducir un texto, a las palabras o sus relaciones, mediante las apariciones conjuntas de las palabras que componen ese texto. Este tipo de análisis permite identificar áreas emergentes de conocimiento, composición temática de líneas de investigación, estructura temática de un proyecto investigativo, etc. (Buzydowski *et al.*, 2002; Fattori *et al.*, 2003; Gil Kim *et al.*, 2008).

Para analizar texto libre dentro de los registros bibliográficos o documentos a texto completo un proceder muy utilizado es establecer relaciones entre dos palabras, contando las veces que estas aparecen juntas en la misma oración o dentro del texto del documento (Callon *et al.*, 1983; Yuen-Hsien *et al.*, 2007). De esta forma se pueden definir las relaciones entre palabras y su grado de intensidad.

Para elaborar los mapas de coocurrencia se parte del principio del valor de las coocurrencias entre dos

agregados, cualesquiera deben ser considerados como el inverso de la distancia (similitud) existente entre esos mismos agregados en una representación bidimensional. Esta correlación negativa entre similitud y distancia permite utilizar diversos procedimientos matemáticos para calcular las coordenadas de los agregados en el mapa a partir de sus distancias. Y permite a su vez, mediante la utilización de una metáfora espacial (Shneiderman & Plaisant, 2006) representar las relaciones existentes entre ellas. Estos mapas no son un objetivo en sí mismo sino el vehículo mediante el cual se recuperan y representan estructuras de conocimiento. Los mapas de relaciones de palabras sirven de herramienta para expresar de forma visual los conglomerados que se establecen entre los contenidos de los documentos, e incluso su presencia y combinación en diferentes campos de un registro bibliográfico (Braam *et al.*, 1991). A partir de estas premisas, este estudio pretende analizar el estado del arte del dominio que se analiza para conocer su estructura de contenidos, y además aplicar las técnicas de análisis de contenido a campos específicos y poco analizados en los estudios de patentes (título, resumen y reivindicaciones), no se encontró dentro de la bibliografía consultada estudios anteriores que aborden de igual manera este proceder.

El objetivo de esta investigación es proponer el uso de la clasificación internacional de patentes y la coocurrencia de palabras en determinados campos del documento de patente como unidad de análisis y medida para el análisis de contenido del dominio tecnológico *vegetable oil combustion*.

Este caso de estudio forma parte de un proyecto de investigación de la empresa Vidrio Artesanal Xaquixe S. de R.L. M.I.ART., con nombre comercial Studio Xaquixe, fundada en el año 2002 como una de las primeras fábricas de vidrio artesanal en México. Tiene como objeto crear piezas artesanales con diseños originales para lograr un impacto en la apreciación del vidrio como forma de arte. Durante los últimos años, la empresa se ha dedicado a mejorar el diseño de su horno, se examina la utilización de nuevos materiales con el fin de construir hornos con una mayor eficiencia y que hagan uso de combustibles diferentes a los no renovables (de origen fósil); de tal manera, que se disminuya el impacto negativo al medio ambiente por el uso de este tipo de hornos en el proceso

de fabricación de los productos. A partir de esta necesidad, el presente estudio tiene el propósito de conocer las tecnologías que existen en el dominio *vegetable oil combustion* e identificar si existe alguna que utilice el aceite vegetal para la combustión interna en hornos. Para alcanzar este objetivo se propone analizar la clasificación temática de las patentes de este dominio tecnológico y realizar un análisis de contenido en diferentes campos del documento de patente: título, resumen y reivindicaciones. Los resultados del estudio apoyarán significativamente el desarrollo de nuevas tecnologías que optimicen los trabajos artesanales con vidrio, como forma de arte.

Métodos y Procedimientos

La metodología utilizada se describe mediante la fuente de datos, estrategia de búsqueda, herramienta para el procesamiento de datos y técnicas para la representación y visualización de la información del dominio.

Fuente de datos: Oficina de Patentes y Marcas de Estados Unidos (USPTO - *United States Patent and Trademark*) se utiliza específicamente como fuente de datos la Base de Datos (BD) Patent Grant.

Estrategia de búsqueda: `abst/((“vegetable oil” OR “vegetable oils” OR “cooking oil” OR “cooking oils”) AND (burn$ OR Combust$) ANDNOT (candle OR lamp)) OR aclm/((“vegetable oil” OR “vegetable oils” OR “cooking oil” OR “cooking oils”) AND (burn$ OR Combust$) ANDNOT (candle OR lamp))`

Procesamiento de los datos: *Software proINTEC* para la descarga, normalización, procesamiento, análisis y visualización de los datos.

Técnicas: Se utiliza como técnica de visualización las redes presentadas en forma de nodos o vértices, donde la estructura semántica se encuentra definida por los enlaces que conectan dichos nodos. La visualización mediante redes hace posible representar las relaciones en red, cruzando variables para encontrar comportamientos o patrones dentro del contenido de las patentes.

Para una mejor representación de la información en el análisis e interpretación de los mapas tecnológicos se utiliza el algoritmo Pathfinder porque permite mostrar

sólo los enlaces más relevantes de las redes mapeadas. Se usa además como técnica de poda, la delimitación numérica de la intensidad de las relaciones en aquellos nodos en que su relación es débil. Y para el posicionamiento de los nodos en la visualización del dominio se emplea el algoritmo de representación espacial Spring Embedded, se utiliza el Netdraw para manipular los gráficos en las representaciones visuales.

Resultados y Discusión

Análisis: indicadores simples

Serie Temporal: El dominio *vegetable oil combustion* está compuesto por un total de 211 patentes concedidas por la USPTO entre 1976 y el 10 de diciembre del año 2013. El análisis de la razón de variación de la producción por años indica que no hay estabilidad en el crecimiento a lo largo del período de tiempo estudiado.

Titulares: Son 103 los propietarios de las patentes relacionadas con *vegetable oil combustion*. Alcanzan el mayor liderazgo los titulares: Paloma Industries, Ltd. en Nagoya, Japón; The Frymaster Corporation en Shreveport, Estados Unidos; Phillips Petroleum Company en Bartesville Estados Unidos y Pitco Frialator, Inc. de Estados Unidos con sede en Concord. Le siguen Gas Research Institute de Chicago, The Lubrizol Corporation en Wickliffe y Heat and Control, Inc. de Hayward todas de Estados Unidos, así como Elf Antar France del estado de Courbevoie en Francia, cada uno con 4 patentes. Otro segmento muy pequeño de titulares tiene 3 patentes (Ethyl Corporation en Richmond y Philip Morris Incorporated en New York, ambos de Estados Unidos); seguidos de un bloque de 16 titulares con 2 patentes cada uno. El resto de los titulares tienen una sola patente.

Inventores: Hay 334 inventores en las patentes concedidas por USPTO entre 1976 y el 2013 sobre el dominio *vegetable oil combustion*.

La mayor productividad inventiva de este dominio recae en la persona de Ejiri; Susumu de Japón, inventor con mayor desempeño tecnológico en esta temática al ostentar 7 patentes en el periodo que se estudia. Le sigue, *Rice George McNair de Shreveport* en

Estados Unidos con 6 invenciones, y por *Yokoyama Nobuyoshi* de Japón con 5 patentes. Se encuentran además otros inventores con relativa productividad respecto a la mayoría, como es el caso de: Kimura Makoto de Tokio, Sank; Gerald W., Benson; Clark K., Caridis Andrew A. y Moore Lewis Frank todos de Estados Unidos, cada uno con 4 patentes.

Este grupo de 8 inventores es el núcleo de investigación e innovación más fuerte de todo el dominio, constituyen el principal frente de investigación de frontera en la temática *vegetable oil combustion*.

Otro grupo de aproximadamente 55 inventores sólo tienen 2 patentes durante el periodo estudiado, mientras el resto de los inventores le han concedido tan solo una patente sobre este tema entre 1976 y el 2013.

Clasificación Internacional de Patentes (CIP): dentro de las normativas que establece la (Organización Mundial de la Propiedad Intelectual, 2014) se encuentra que toda invención debe acreditarse a una o varias clasificaciones en el campo número 51 del documento de patente. La Clasificación Internacional de Patentes identifica y describe las áreas del conocimiento tecnológico a la cual pertenece cada invención.

La Clasificación Internacional de Patentes abarca todo el conjunto de conocimientos que pueden considerarse dentro del ámbito científico-tecnológico, y de forma específica, los que trabajan las patentes de invención. Uno de los autores que más trabaja el tema de las clasificaciones de patentes es Leydesdorff, quién incluso considera las clases de la clasificación como equivalentes de las revistas (Leydesdorff, 2008a; 2008b).

El conocimiento tecnológico en la CIP está dividido en ocho secciones, constituye la Sección el nivel jerárquico más alto de esta clasificación. Cada sección de la CIP tiene un título (compuesto por una o varias palabras) y un símbolo (una letra mayúscula del alfabeto romano). El título de cada sección va seguido de un resumen de los títulos de sus subdivisiones principales.

El dominio *vegetable oil combustion* aborda dentro de las reivindicaciones que protegen sus patentes 7 áreas del conocimiento de la técnica.

El mayor número de patentes están clasificadas en la *sección A* relacionada con la temática *necesidades corrientes de la vida*, se infiere que las aplicaciones

tecnológicas de este dominio tienen un alto uso en variadas esferas de la vida. Esta *sección* técnica tiene además patentes en casi todos los años bajo estudio, alcanzando la mayor sostenibilidad temporal dentro del dominio.

Le sigue la *sección C* (Química) con 72 patentes concedidas en casi todos los años bajo estudio, así como la *sección F* (Mecánica. Iluminación. Calefacción. Armamento. Voladura) con 56 patentes. Ambas clasificaciones tienen una fuerte presencia investigativa en casi todos los años del periodo que se analiza.

La clasificación que aparece como emergente es la *sección G* (Física), al tener el mayor número de sus patentes concedidas en los últimos años; algunas patentes concedidas en este siglo han protegido conocimientos relacionados con la *sección G*.

El análisis por el *segundo nivel jerárquico* de la CIP reveló que las invenciones sobre *vegetable oil combustion* se ubican en más de una veintena de clases temáticas.

El análisis de las clases técnicas por años proyectó la siguiente composición temática:

- Las clasificaciones más trabajadas en el periodo comprendido entre los años 1976 y el año 1986 son las siguientes: A23, A24, A47, A61, A62, A63, B1, B41, B65, C1, C10, C6, C8, C9, F1, F16, F2, F21, F23, F24, F28, G1, G5, G8, G9, H1 y H5.

- Las clases temáticas trabajadas después del año 2000 son: A1, A47, A61, B1, B32, B60, B9, C10, C23, C6, C7, C8, C9, F1, F16, F2, F23, F24, G1, G5, G6 y la H5.

- Las clases presentes en las patentes del año 2013 son: A61, B60, B9, C10, C7, F23, G5 y G6.

El análisis por el *tercer nivel jerárquico* de la CIP reveló que las patentes sobre *vegetable oil combustion* ubican el conocimiento tecnológico de sus reivindicaciones en un amplio número de sub clases temáticas.

- Las temáticas trabajadas en el dominio *vegetable oil combustion* se concentran en 27 subclases después del año 2000.

- Las subclases temáticas trabajadas en el presente siglo son: A1N, A47J, A61K, A61L, A61Q, B1D, 1J, B32B, B60L, B60P, B65D, C10J, C10L, C1B, C1F, C6C, C7C, C7D, C8J, C8K, C8L, F23B, F23D, F23G, G5D, G6F y G6G.

- Y las subclases temáticas que abordan las patentes del año 2013 son: A61K, A61L, B60L, B9B, C10L, C10M, C7D, F23B, F23D, F23G, G5D, G6F y G6G.

Relaciones Temáticas

El dominio *vegetable oil combustion* tiene una estructura de relaciones temáticas compuesta por 7 secciones del conocimiento tecnológico. Las relaciones interdisciplinarias más fuertes ocurren entre la sección técnica de Mecánica (F) y la (A) que representa la aplicación de las tecnologías o procesos que se desarrollan. También existen relaciones triangulares entre (A, F, B)-(H, A, F)-(H, A, G), etc. En particular, a esta investigación le interesa destacar la triangulación establecida entre la sección F (Mecánica), la sección A (Necesidades corrientes de la vida) y la sección que se ocupa de Técnicas Industriales Diversas (B).

Las clasificaciones conjuntas que tienen mayor intensidad en el *segundo nivel jerárquico de la CIP* del dominio son a partir de la clase A47 con otras clases; así como la triangulación presente entre A47-F23-A23. En este último caso, se manifiesta la relevancia de las investigaciones relacionadas con aparatos de combustión. Procesos de combustión (F23) dentro de este dominio.

Al profundizar el análisis hasta el *último nivel jerárquico de clasificación* se obtuvieron las principales clasificaciones conjuntas de este dominio. En los diferentes clúster temáticos que muestra este análisis aparecen delimitados los diferentes nodos que existen incluso, dentro de una misma sección del conocimiento tecnológico de este dominio.

Las relaciones más intensas se manifiestan entre las clasificaciones A47J0037/120000 y la A47J0027, y éstas con la clasificación F23C0015, constituyen la relación de triangulación más intensa e interesante de todo el dominio. Esta relación indica que existe un gran número de patentes concedidas que han protegido invenciones relacionadas dentro del área de Mecánica con procedimientos o aparatos de combustión que utilizan combustibles fluidos o fluidizados, principalmente aparatos en los que la combustión tiene lugar en pulsos influenciados por resonancia acústica en una masa de gas (F23C0015).

Otra fuerte presencia recae en la sección C donde los subgrupos que mayor relación están relacionados con combustibles de líquidos carbonosos, dentro de ellos los basados esencialmente en componentes constituidos únicamente por carbono, hidrógeno y oxígeno; así como los que contienen aditivos, y compuestos orgánicos, etc.

A esta investigación le interesan las relaciones que se establecen principalmente con el subgrupo que abarca combustibles no previstos en otros lugares. Adición de sustancias a los combustibles o al fuego para reducir el humo o depósitos indeseables, o para facilitar la eliminación del hollín. Generadores de fuego (C10L0001/180000), subgrupo temático que presenta el mayor núcleo de relaciones dentro del área de la Química.

Área Temática de los Inventores: El análisis del frente temático de los inventores de este dominio manifestó que los investigadores más productivos están patentando sus invenciones en el área de Mecánica (F). Se conoció que otro gran número de investigadores trabajan en patentes relacionadas con la sección A (201 inventores), C (167 inventores), F (117 inventores), B (41 inventores), H (24 inventores) y E (1 inventor).

Área Temática de los Titulares: Al analizar el comportamiento de los titulares por secciones de la CIP se conoció la estructura temática de los signatarios, así como las relaciones temáticas entre ellos (Figura 1). Se identificó que:

- *Paloma Industries, Ltd.* trabaja principalmente en 3 secciones de la CIP (mecánica-F, física-G y necesidades corrientes de la vida-A);

- *Gas Research Institute* trabaja invenciones relacionadas con la sección de mecánica -F y necesidades corrientes de la vida-A;

- *The Frymaster Corporation* tiene invenciones relacionadas con física-G y necesidades corrientes de la vida-A;

- *Philip Morris Incorporated* trabaja invenciones relacionadas con la sección de necesidades corrientes de la vida-A y química. metalurgia-C);

- *Thermo Electron Corporation* tiene invenciones relacionadas con necesidades corrientes de la vida-A y técnicas industriales diversas-B.

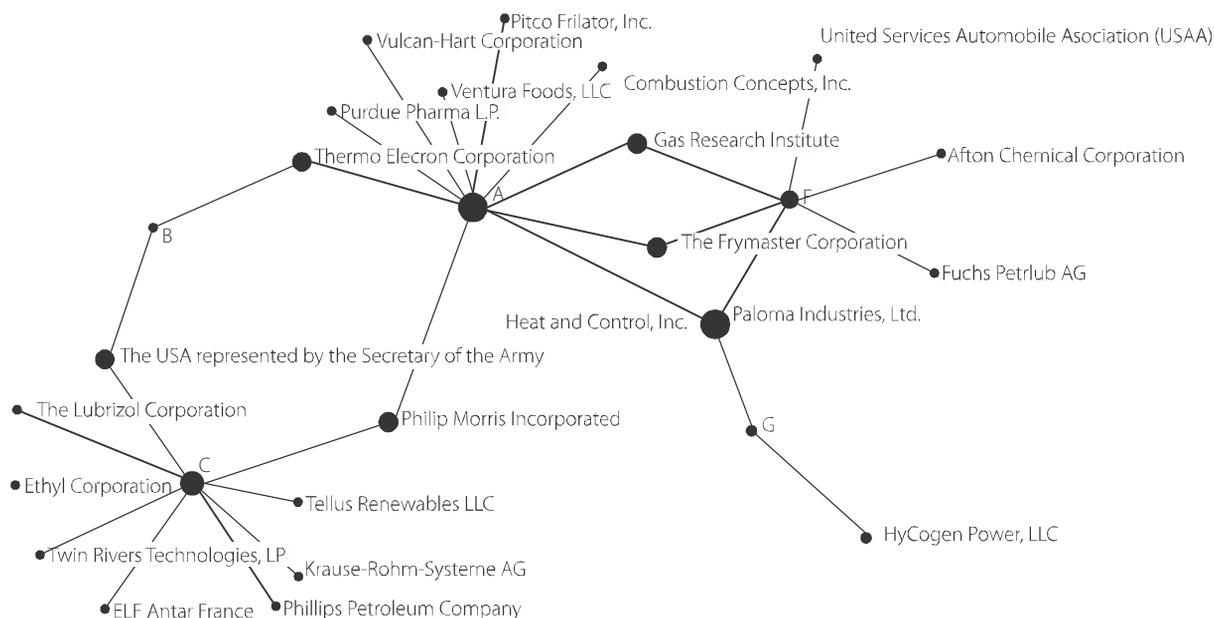


Figura 1. Titulares por sección Clasificación Internacional de Patentes (CIP) mayores que 1.

Fuente: *Software* proINTEC.

Lo más significativo de este análisis es encontrar dentro de esta sección F a los principales titulares, o sea a los de mayor productividad. Aquí también escalan las primeras posiciones, tal y como ocurre con el comportamiento de los inventores por secciones temáticas.

Al analizar el comportamiento temporal de la sección F se obtuvo que las patentes concedidas sobre esta temática están presentes desde 1976, existe el mayor incremento en investigaciones de esta temática en el año 1990 (5 patentes concedidas). A partir de ese año, se han mantenido las investigaciones relacionadas con la sección de Mecánica dentro *vegetable oil combustion*.

Análisis contenido (Co-word)

Para realizar el análisis del contenido de las patentes de este dominio se aplicó la técnica de *coword* en diferentes campos del documento de patente. En este caso, los campos del registro de patente que tienen un alto contenido de información importante a relacionar son: título, resumen y reivindicaciones.

Las reivindicaciones definen el objeto por el que se solicita protección. Su redacción debe ser clara y concisa, argumentándose en la descripción. El contenido de las reivindicaciones es de esencial interés puesto que son ellas las que determinan la extensión de la protección conferida por la patente y el alcance del objeto que se protege. En las reivindicaciones se encuentra el objeto de invención y están contenidas las principales palabras que describen la nueva tecnología patentada, así como en los campos título y resumen.

El primer estudio (Figura 2) analizó los campos título, resumen y reivindicaciones de los documentos de patentes del dominio; en este caso, el enlace máximo es de 71 y el mínimo de 7.

La parte superior del mapa muestra que existen tecnologías relacionadas con este tema.

La relación más intensa se establece entre Oils y Vegetables, y este último término con minerales, ácidos, etc. revela de alguna manera que mantienen relación con la utilización de combustible de origen vegetal, o sea, esto puede estar relacionado con procesos o tecnologías de obtención de combustibles a partir del aceite o grasas vegetales.

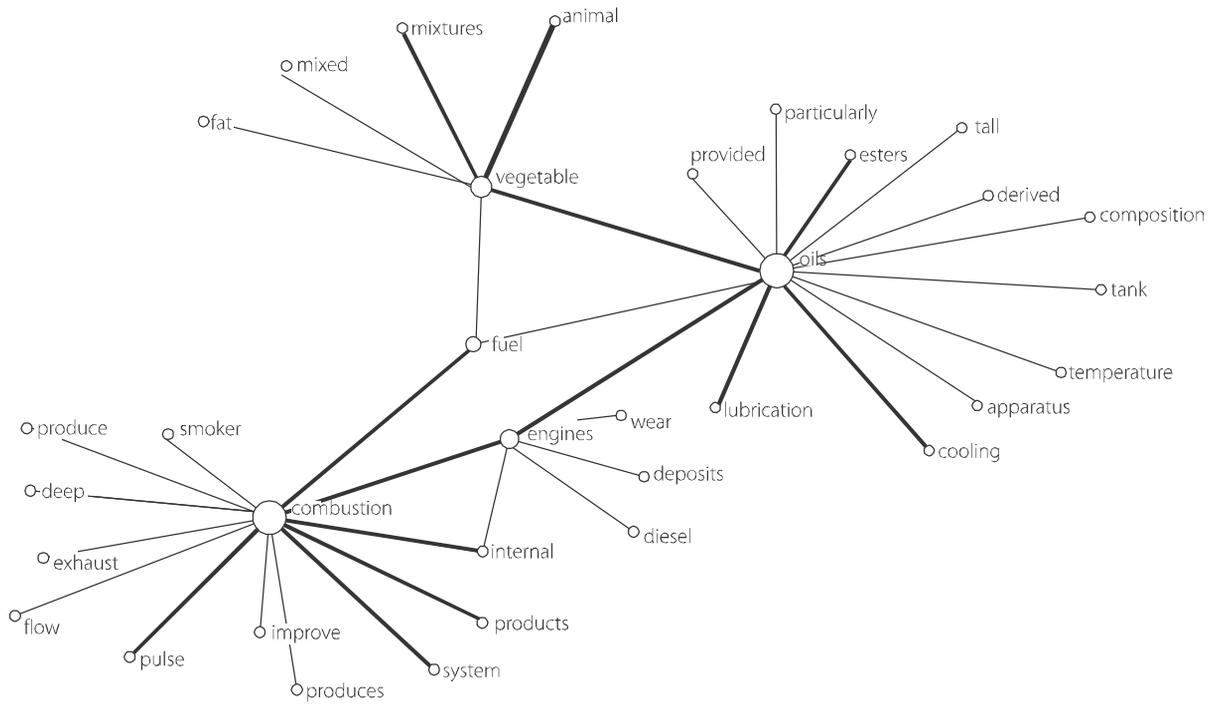


Figura 4. Cword en el campo título y resumen. Personalizado.

Fuente: Software proINTEC.

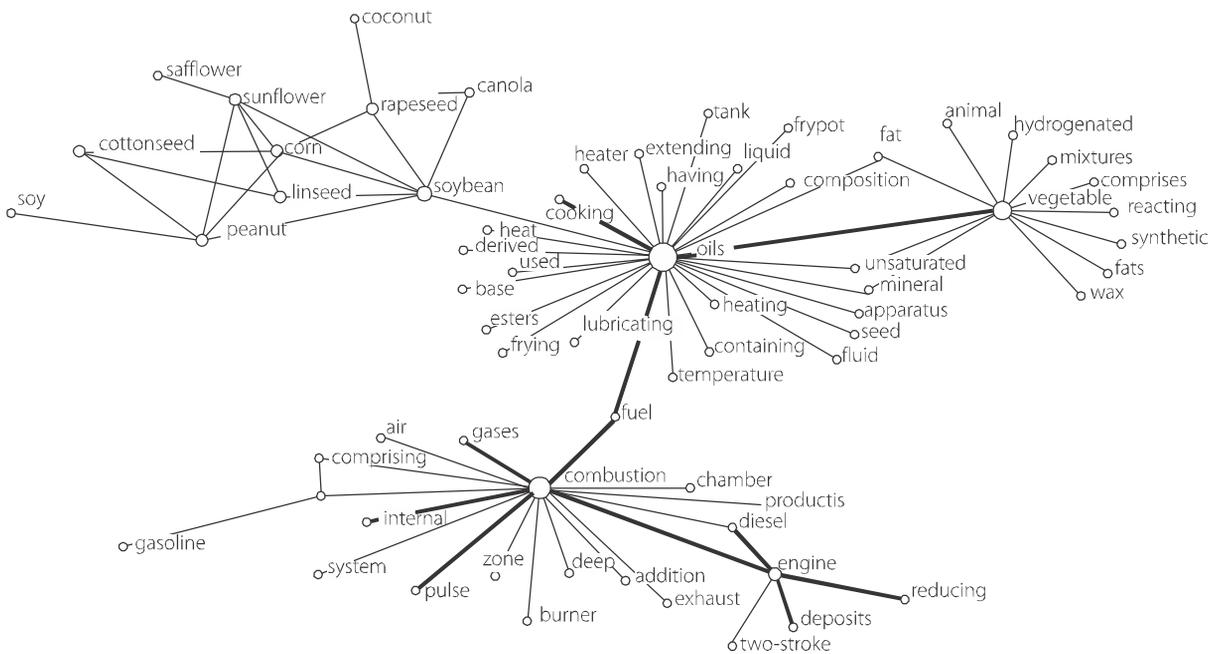


Figura 5. Cword en el campo título, resumen y reivindicaciones. Personalizado.

Fuente: Software proINTEC.

El mapa representa la cadena de relaciones de esta tecnología, o sea el uso de las semillas para obtener aceites y el uso de los aceites como combustión, etc. e indica que no se usan de manera directa en la combustión. Y por otra parte, todas las palabras asociadas a quemadores están relacionadas con cocinas y aceites, esto pudiera significar que desarrollan estos quemadores para diferentes usos, incluyendo la cocina.

Del mapa resultante se obtuvo evidencias claras de las relaciones tecnológicas que existen en la obtención a partir de diferentes semillas de aceites vegetales que permitan generar combustibles para máquinas de combustión interna.

Conclusión

El dominio *vegetable oil combustion* está compuesto por un total de 211 patentes concedidas por la USPTO entre 1976 y el 10 de diciembre del año 2013. La mayor productividad sostenida se observa a partir del año 2011, aunque la razón de variación de la producción por años indica que este dominio ha estado fluctuando su desarrollo durante todo el periodo que se analiza.

Existen 103 propietarios en el dominio *vegetable oil combustion*. Alcanzan el mayor liderazgo los titulares: *Paloma Industries, Ltd.* de Japón y *The Frymaster Corporation*; *Phillips Petroleum Company* y *Pitco Frialator, Inc.* de Estados Unidos.

El país que mayor liderazgo ostenta introduciendo nuevos procesos y tecnologías en el mercado internacional sobre *vegetable oil combustion* y el que mayor poderío tecnológico tiene es Estados Unidos (72 invenciones). Considerado el país que mayor infraestructura tecnológica y empresarial exhibe en esta temática. Le siguen Japón (12 titulares), Alemania (6 titulares), Francia (4 titulares), Canadá (3 titulares) y China con 2 titulares. El resto de los países sólo tienen un titular que trabaja en este dominio.

Hay 334 inventores en las patentes concedidas por USPTO entre 1976 y el 2013 sobre el dominio *vegetable oil combustion*. La mayor productividad inventiva y desempeño tecnológico lo exhibe el investigador Ejiri; Susumu de Japón (7 patentes), seguido

por Rice George McNair de Estados Unidos (6 invenciones) y Yokoyama Nobuyoshi de Japón (5 patentes). Le acompañan otro grupo de inventores con relativa productividad respecto a la mayoría: Kimura Makoto, Sank; Gerald W., Benson; Clark K., Caridis Andrew A. y Moore Lewis Frank, cada uno con 4 patentes. Este conjunto de 8 inventores son el núcleo de investigación e innovación más fuerte de todo el dominio, constituyen el principal frente de investigación de frontera en la temática *vegetable oil combustion*.

Algunos de los inventores emergentes que aparecen en las investigaciones más recientes del dominio son: Germanaud; Laurent (3 patentes) y Sackler; Richard S, Sagar; Christopher Lee, Nam; Jong Hyun con 2 patentes, etc.

Los investigadores más productivos del dominio patentan principalmente en el área de Mecánica, Necesidades corrientes de la vida y en menor alcance en la sección de Química. Así como los titulares más productivos son los que trabajan fundamentalmente en la sección de Mecánica.

El análisis de colaboración por titulares reveló que existe un grupo de organizaciones que trabajan de manera conjunta, pero los titulares que aparecen en cotitulación no son de los más productivos del dominio. Los titulares de mayor productividad son los únicos propietarios de las tecnologías que desarrollan y no trabajan sus investigaciones en colaboración.

El dominio *vegetable oil combustion* aborda 7 áreas del conocimiento de la técnica. El mayor número de patentes se clasifican en la sección A relacionada con temáticas sobre necesidades corrientes de la vida, además es la sección que le han concedido patentes en casi todos los años bajo estudio. Esto manifiesta que las aplicaciones de las tecnologías de este dominio tienen un alto uso en variadas esferas de la vida. Le sigue la sección C de química y la sección F de mecánica, con una fuerte presencia investigativa en casi todos los años del periodo que se analiza.

Las relaciones interdisciplinarias más fuertes que ocurren en este dominio se establecen entre la sección técnica de Mecánica (F) y la sección A, ellas representan la aplicación de las tecnologías o procesos que se

desarrollan. La triangulación establecida entre la sección de Mecánica (F), la sección A y la sección de Técnicas Industriales Diversas (B) manifiesta la relevancia de las investigaciones relacionadas con aparatos de combustión y procesos de combustión dentro de este dominio.

El análisis de relaciones temáticas en sus niveles jerárquicos muestra la presencia de invenciones relacionadas dentro del área de Mecánica con procedimientos o aparatos de combustión que utilizan combustibles fluidos, principalmente aparatos en los que la combustión tiene lugar en pulsos influenciados por resonancia acústica en una masa de gas (F23C). Una fuerte presencia recae en Química donde los subgrupos que mayor relación tienen/están relacionadas con combustibles de carbonosos líquidos, dentro de ellos los basados esencialmente en componentes constituidos únicamente por carbono, hidrógeno y oxígeno; así como los que contienen aditivos, y compuestos orgánicos, etc.

Otras relaciones interesantes se establecen con el subgrupo que abarca combustibles no previstos, adición de sustancias a los combustibles o al fuego para reducir el humo o depósitos indeseables, o para facilitar la eliminación del hollín, generadores de fuego, etc. Subgrupo temático que además presenta el mayor núcleo de relaciones dentro del área química.

Del análisis de contenido (co-Word) se obtuvo evidencias claras de la obtención de aceites vegetales a partir de diferentes semillas para la generación de combustibles para máquinas de combustión interna.

Oils tiene una fuerte asociación con vegetable, mixtures, animal, esters, apparatus, fuel, lubrication y se puede deducir que de alguna manera se desarrollan tecnologías para utilizar aceite vegetal como combustible en máquinas de combustión interna, así como tecnologías que utilizan aceites vegetales como lubricantes en máquinas de combustión interna.

Referencias

- Basberg, B. Foreign patenting in the U.S. as a technology indicator: The case of Norway. *Research Policy*, v.12, n.4, p.227-237, 1983.
- Basberg, B. Technological change in the Norwegian whaling industry: A case-study in the use of patent-statistics as a technology indicator. *Research Policy*, v.11, n.3, p.163-171, 1982.
- Braam, R.R.; Moed, H. F.; Van R.; Anthony F. J. Mapping of science by combined co-citation and word analysis. I: structural aspects. *Journal of the American Society for Information Science*, v.42, n.4, p.233-251, 1991.
- Buzdylowski, J.; White, H. D.; Lin, X. Term co-occurrence analysis as an interface for digital libraries. In: Bomer, K. & Chen, C. (Ed.). *Visual interfaces to digital libraries*. Berlin: Springer, 2002, p.133-144. (Lecture Notes in Computer Science Series, v.2539).
- Callon, et al. From translation to problematic networks an introduction to co-word analysis. *Social Science Information Sur les Sciences Sociales*, v.22, n.2, p.191-235, 1983.
- Díaz-Pérez, M.; Moya-Anegón, F. El análisis de patentes como estrategia oportuna para la toma de decisiones innovadoras. *El Profesional de la Información*, v.17, n.3, 2008.
- Fattori, M.; Pedrazzi, G.; Turra, R. Text mining applied to patent mapping: A practical business case. *World Patent Information*, v. 25, n.4, p.335-342, 2003.
- Gil Kim, Y.; Hwan Suh, J.; Chan Park, S. Visualization of patent analysis for emerging technology. *Expert Systems with Applications*, v.34, n.3, p.1804-1812, 2008.
- Grupp, H.; Schmoch, U.; Kuntze, U. Patents as potential indicators of the utility of ec research programmes. *Scientometrics*, v.21, n.3, p. 417-445, 1991.
- Klitkou, A.; Gulbrandsen, M. The relationship between academic patenting and scientific publishing in Norway. *Scientometrics*, v.82, n.1, p.93-108, 2010.
- Leydesdorff, L. The delineation of nanoscience and nanotechnology in terms of journals and patents: A most recent update. *Scientometrics*, v.76, n.1, p.159-167, 2008a.
- Leydesdorff, L. Patent classifications as indicators of Intellectual organization. *Journal of the American Society for Information Science and Technology*, v.59, n.10, p.1582-1597, 2008b.
- Leydesdorff, L.; Heimeriks, G. The self organization of the European Information Society: The case of "biotechnology". *Journal of the American Society for information Science and Technology*, v.52, n.14, p.1262-1274, 2001.
- Narin, F. Patent bibliometrics. *Scientometrics*, v.30, n.1, p.147-155, 1994.
- Narin, F.; Olivastro, D. Technology indicators based on patents and patent citations. In: Van Raan, A.F.J. (Ed.). *Handbook of quantitative studies of science and technology*. Amsterdam: Elsevier Publishers, 1988.
- Narin, F. Patents as indicators for the evaluation of industrial research output. *Scientometrics*, v.34, n.3, p.489-496, 1995.
- Organización Mundial de la Propiedad Intelectual. *Clasificación internacional de patentes: guía*, 2014.

Disponível em: <<http://www.wipo.int/pct/es/>>. Acesso em: 10 jun. 2013.

Pao-Long, C.; Chao-Chan, W.; Hoang-Jyh, L. Using patent analyses to monitor the technological trends in an emerging field of technology: a case of carbon nanotube field emission display. *Scientometrics*, v.82, n.1, p.5-19, 2010.

Pavitt, K. Patent statistics as indicators of innovative activities: Possibilities and problems. *Scientometrics*, v.7 n.1-2, p.77-99, 1985.

Pavitt, K. Uses and abuses of patent statistics. In: Van Raan, A.F.J. (Ed.). *Handbook of Quantitative studies of science and technology*. Amsterdam: Elsevier Publishers, 1988. p.509-536.

Rip, A.; Courtial, J.P. Co-word maps of biotechnology- an example of cognitive scientometrics. *Scientometrics*, v.6, n.6, p.381-400, 1984.

Shneiderman, B.; Plaisant, C. Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In: Workshop on Beyond Time and Errors: Novel Evaluation Methods For information Visualization BELIV'06, 2006, New York. *Proceedings...* New York: ACM Press, 2006. p.1-7.

Spinak, E. Indicadores cienciométricos. *Ciência da informação*, v.27, n.2, p.141-148, 1998.

Yuen-Hsien T.; Chi-Jen, L.; Yu-I L. Text mining techniques for patent analysis. *Information Processing and Management*, v.43, n.5, p.1216-1247, 2007.

Estudio de patentes sobre tecnologías para tratamiento de agua y el agua residual¹

A study of patenting on water and wastewater treatment technologies

Odonel GONZÁLEZ-CABRERA²

Taimi RUIZ-PÉREZ³

Mirelys CLARO-PÉREZ²

Nancy María PÉREZ-PINO²

Grisel PÉREZ-GÁLVEZ²

Lisbey COLLAZO-ALFONSO²

Resumen

Se presenta un estudio de patentes sobre la temática tratamiento de agua y agua residual. Se localizan y descargan más de 2.500 documentos de patentes relacionados con la temática. Estos se procesan en *software* al efecto y herramientas de procesamiento. Se extraen indicadores de patentes para esta tecnología con alto valor para actividades comerciales y de investigación, tales como tecnologías emergentes sobre tratamiento de agua, países, entidades e inventores líderes a nivel mundial según el resultado de la tecnología patentada. Se presenta la evolución de las tecnologías de tratamiento del agua y el agua residual en los últimos veinte años a nivel mundial según el registro de las patentes. La entrega de todo el material se realiza en una herramienta de visualización que logra hacer más atractivo el servicio, tener un producto más amigable y cómodo para el uso, constituyendo una información de alto nivel de actualización y eficacia para la toma de decisiones de una empresa de proyectos e investigaciones hidráulicas.

Palabras clave: Documento de patente. Indicadores de patentes. Tratamiento de agua.

Abstract

The aim of the study is to discuss the patenting of water and wastewater treatment technologies. More than 2.500 patent documents related to the topic were found and downloaded. The documents were analyzed by a software and processing tools. Patent indicators for the technology with high-value business and research were determined such as water treatment, countries, institutions and worldwide leader inventors according to the patented technology. The worldwide development of water and wastewater treatment technologies over the past twenty years is presented according to the record of patents. The delivery of all materials is done in a visualization tool that manages to make service more attractive, it is a friendlier and more comfortable product to use that offers a high-level update and effectiveness of information for decision making in a hydraulic project and research business.

Keywords: Patent document. Patent indicators. Water treatment.

¹ Trabajo presentado en el VII Seminario Internacional sobre Estudios Cuantitativos y Cualitativos de la Ciencia y la Tecnología "Prof. Gilberto Sotolongo Aguilar en XIII Congreso Internacional de Información - INFO'2014. Habana, Cuba.

² Centro de Información y Gestión Tecnológica de Villa Clara. Marta Abreu #55 entre Villuendas y Sayas, 50100, Santa Clara, Cuba. Correspondencia a nombre de/Correspondence to: O. GONZÁLEZ-CABRERA. E-mail: <ogonzalez@ciget.vcl.cu>.

³ Hospital Docente Gineco-obstétrico "Mariana Grajales". Santa Clara, Cuba.

Recibido el día 21/5/2014 y aceptado para su publicación el 12/9/2014.

Introducción

El agua es un recurso fundamental para la vida tanto de los seres humanos como del resto de las especies de la tierra. Con un planeta cada vez más poblado y el cambio climático haciendo estragos en muchas partes del mundo, la gestión, conservación y acceso al agua en buenas condiciones es cada vez más difícil sobre todo en los países en desarrollo (Alemany, 2013).

El 59% del consumo total de agua en los países desarrollados se destina a uso industrial, el 30% a consumo agrícola y un 11% a gasto doméstico, según se constata en el primer informe de Naciones Unidas sobre el desarrollo de los recursos hídricos del mundo (Fernández-Alba *et al.*, 2006). Conscientes de la problemática en torno a este recurso, Naciones Unidas trabaja para conseguir que todos los ciudadanos del mundo puedan tener acceso al agua potable y condiciones de saneamiento dignas. Para ello el período comprendido entre 2005 y 2015 se declaró el Decenio Internacional para la Acción "El agua, fuente de vida". Dentro de las líneas de actuación de este proyecto, la Asamblea General de la Organización de Naciones Unidas (ONU) aprobó que el 2013 fuera el "Año Internacional de la Cooperación en la Esfera del Agua", en su resolución 65/154 con el objetivo de sensibilizar tanto sobre el potencial de una mayor cooperación como sobre los desafíos que afronta la gestión del agua debido a una mayor demanda y lo que ello implica para el acceso, la asignación y los servicios. El Año pondrá en valor la historia de las iniciativas exitosas de cooperación en materia de agua e identificará los temas emergentes y más urgentes.

El director general de la Organización Mundial de la Salud (OMS) Dr. Lee Jong-Wook plantea que el agua y el saneamiento son unos de los principales motores de la salud pública. Suelo referirme a ellos como "Salud 101", lo que significa que en cuanto se pueda garantizar el acceso al agua salubre y a instalaciones sanitarias adecuadas para todos, independientemente de la diferencia de sus condiciones de vida, se habrá ganado una importante batalla contra todo tipo de enfermedades (Wikipedia 2013).

Para poder contar con estos elevados fines debemos contar con los conocimientos de las tecnologías existentes para el tratamiento del agua y el agua residual, saber quiénes las han generado (inventores), quienes son sus dueños (titulares de las patentes), en que países están localizadas estas personas, empresas o centros de investigación (tema bloqueo y otros), nivel de obsolescencia de la tecnología y otros.

El tratamiento de agua y aguas residuales consiste en una serie de procesos físicos, químicos y biológicos que tienen como fin eliminar los contaminantes físicos, químicos y biológicos presentes en el agua para el uso humano, así como los efluentes de este. La tesis fundamental para el control de la polución por aguas residuales ha sido tratar las aguas residuales en plantas de tratamiento que hagan parte del proceso de remoción de los contaminantes y dejar que la naturaleza lo complete en el cuerpo receptor. Para ello, el nivel de tratamiento requerido es función de la capacidad de auto purificación natural del cuerpo receptor.

Por otro lado podemos exponer que existe un tipo de información tecnológica especial contenido en documentos llamados patentes de invención a los cuales por disímiles causas no se les presta la atención debida a la hora de resolver problemáticas como podría ser el tratamiento y potabilización de aguas. Las patentes constituyen una forma de proteger los inventos que desarrollan las empresas, instituciones o personas y que cumplan con los requerimientos de novedad mundial, actividad inventiva y aplicabilidad tecnológica; esta protección la otorga una oficina estatal y consiste en otorgarle el derecho de monopolio para su venta, producción, ofrecimiento de licencia, etc. Pero a la vez obliga al titular a revelar todos los detalles técnicos de su invención y es precisamente este hecho el que aprovechamos para extraer la información precisa y actualizada (estos documentos se publican por las oficinas de patentes en un plazo posterior solo de 18 meses posterior a su presentación) de cualquier rama de la tecnología. También se realiza el procesamiento de la totalidad de documentos recuperada estadísticamente para extraer numerosos indicadores de patentes útiles para la toma de decisiones en las más disímiles disciplinas: vigilancia y transferencia tecnológica, búsqueda de potenciales, socios comerciales, encontrar la solución a

problemas que tenemos y que alguien en el mundo ya solucionó, encontrar nichos tecnológicos para investigación, etc. La finalidad declarada del sistema de patentes es fomentar la invención y el progreso técnico facilitando un plazo temporal de exclusividad sobre la invención a cambio de su divulgación. Al ofrecer protección y exclusividad, la patente constituye un instrumento político cuya finalidad es alentar a los inventores a que inviertan en investigación y en el trabajo de innovación posterior que conseguirá que esas invenciones tengan una aplicación práctica (Organización para la Cooperación y el Desarrollo Económicos, 2009).

Aprovechando el poder y facilidades que aporta el desarrollo de las Tecnologías de la Información y las Comunicaciones (TIC), se han desarrollado softwares para convertir la información de patentes en información gráfica clara o mapas, para permitir que el analista tenga una visión de los datos, saquen conclusiones y minimicen los riesgos en la hora tomar decisiones (Keim, 2002). También se ha designado como patinformática, a la ciencia de análisis de información sobre patentes para descubrir relaciones y tendencias que serían difíciles de ver cuando se trabaja con documentos de patentes de uno en uno (Trippe, 2003).

Basándonos en todo lo anterior expuesto realizamos el estudio sobre las tecnologías para el tratamiento y potabilización de agua para la Empresa de investigaciones y proyectos hidráulicos de Villa Clara. Presentando como objetivo principal: Fundamentar la toma de decisiones de una empresa de investigación y proyectos hidráulicos en temáticas de vigilancia tecnológica y transferencia de tecnología sobre tecnologías de tratamiento de agua y agua residual.

Métodos

Se implementan estrategias de búsqueda en la base de datos de patentes europea Esp@cenet (European Patent Office, 2013) bajo la temática de tratamiento de agua y agua residual. La estrategia combina los diferentes campos de búsqueda (palabras claves, fechas, clasificaciones, nombres) para recuperar un número de documentos manejables y sin que se excluyan los

principales resultados a nivel mundial en la temática solicitada. Se trazó la siguiente estrategia de búsqueda:

Palabras claves en título: "Waste water"

Palabras claves en título o resumen: *Treatment*

Fecha de publicación: 2006-2012

Código de la Clasificación Internacional de Patentes (CIP): C02F

C: Química

02F: Tratamiento del agua, agua residual, de alcantarilla o fangos

Cantidad de documentos de patentes: 2826

Base de datos: Esp@cenet

La totalidad de los datos bibliográficos se procesa en software de procesamiento de patentes desarrollado por el Centro de información y Gestión tecnológica de Villa Clara. Se extraen indicadores de patentes. Luego se completan documentos tales como: datos sobre el servicio, estrategias implementadas, breve introducción a la temática en cuestión, comportamiento de los indicadores de patentes, los resúmenes de todas las patentes procesadas y perfiles tecnológicos de las principales entidades generadoras de patentes en la temática. Utilizamos la base de datos (An open..., 2013) que actualmente cuenta con más de 10 millones de documentos de patentes incluidos los registros de las bases de Estados Unidos de América, Australia, Europa y *World Intellectual Property Organization* (WIPO). También nos hemos auxiliado temporalmente de Boliven Pro (Patents, 2013) procesador de patentes on line de Cambridge IP de la Universidad de Cambridge del Reino Unido y los directorios de patentes patentmaps (Wordwide..., 2013) y (Intellectual Property Exchange, 2013). Estas herramientas nos ayudan a de alguna manera validar el resultado que se obtiene con el software. Finalmente con este volumen de información se conforma el boletín en la herramienta de visualización desarrollado por el Centro de información y Gestión tecnológica de Villa Clara, en el cual permite manejar de forma fácil y amigable tan gran volumen de información.

Para la identificación de las tecnologías emergentes nos apoyamos en la CIP. Según la guía de la CIP, una invención se incluye en un grupo de la clasificación según su función o naturaleza intrínseca o

bien por su campo de aplicación. Por tanto, éste es un sistema de clasificación combinado por funciones y aplicaciones en el que prevalece la aplicación. Una patente puede contener diversos objetos técnicos y ser asignada en consecuencia a varias clases CIP; estos códigos de la clasificación se publican en los documentos de patente. De manera global el comportamiento de las tecnologías para el tratamiento de aguas residuales es el siguiente:

En la CIP la clasificación C02F relaciona los documentos de patentes relacionados con tratamiento del agua, agua residual, de alcantarilla o fangos. Seguidamente a estas siglas el número 1 indica tratamientos en general, el número 3 distingue a los tratamientos biológicos y el número 9 hace alusión a tratamientos en varias etapas.

Resultados

Apoyados en las propiedades estadísticas de los datos de patentes que están determinadas por sus características jurídicas y su aplicación práctica en el ámbito económico, puesto que influyen en qué invenciones se protegen, quién las hace, qué información se hace pública (y, por tanto, accesible a los profesionales de la estadística) y qué importancia tienen las patentes para la industria, se extraen numerosos indicadores que nos ayudan en las más disímiles tareas del quehacer de una empresa de investigaciones y proyectos de obras hidráulicas.

Los indicadores basados en patentes constituyen una herramienta única, y en ocasiones la única disponible para seguir el ascenso de las tecnologías emergentes. El primer indicador nos entrega el comportamiento en el tiempo del registro de patentes, visualizando una pendiente creciente en esta tecnología, generándose gran cantidad de patentes cada uno de años siendo objeto de las pesquisas. Desde el año 2007 hasta el 2011 se presentan registros por encima de los 400, especial destaque para el 2010 (467). Este comportamiento muestra el interés elevado por esta temática dado la problemática de la contaminación de las aguas y la necesidad de agua para los más diversos usos.

Seguidamente se identifican a nivel mundial los países líderes en el registro de patentes en esta

tecnología. Este indicador se toma de la fecha de prioridad; porque esta es la primera fecha de presentación de una solicitud de patente en cualquier lugar del mundo para proteger una invención y que por lo general se deposita en la oficina de patentes del país del solicitante. Es la más cercana a la fecha de la invención. Además de la cuestión práctica de este dato, incluye la sigla del país de prioridad.

Las patentes se utilizan para comprobar el rendimiento tecnológico de los países. En comparación con otros indicadores de producción, como pueden ser las publicaciones, las patentes constituyen un indicador más adecuado de las actividades más cercanas al desarrollo tecnológico. Ayudan a rastrear el liderazgo o posicionamiento en un campo o área determinados de la tecnología (por ejemplo, índices de ventajas tecnológicas reveladas) y los cambios que se producen con el tiempo. Como indicadores del rendimiento tecnológico, es decir, del nivel de especialización o fortaleza tecnológica de un país, las patentes ayudan a los gestores de políticas a identificar los puntos fuertes y débiles de sus sistemas de innovación nacionales.

China constituye abrumadoramente el país mayor productor y de mayor registro de las patentes (1.593 para 56,00 % del total), seguido de Japón (463 y 16,00 %), Corea del Sur (338 y 12,00%) y Alemania (106 y 3,75%) entre otros (Figura 1).

Importante es identificar las firmas más destacadas, a este fin tributa un dato contenido en documento de patente llamado titular de la patente, esto no es más que la persona o personas, físicas o jurídicas, a quienes el inventor transfiere jurídicamente en su totalidad o de forma limitada los derechos dimanantes de una patente. Este puede ser un centro de Investigación y Desarrollo; una universidad, una empresa o una persona natural (generalmente el mismo inventor).

Es notable en los últimos años el impacto de las universidades en este particular, siendo verificable recopilando datos de las patentes que han obtenido, las citas que generan, etc. En un número cada vez mayor de países, agencias o ministerios encargados de aportar financiación, estos utilizan el número de patentes para evaluar el rendimiento de las instituciones académicas o de investigadores determinados.

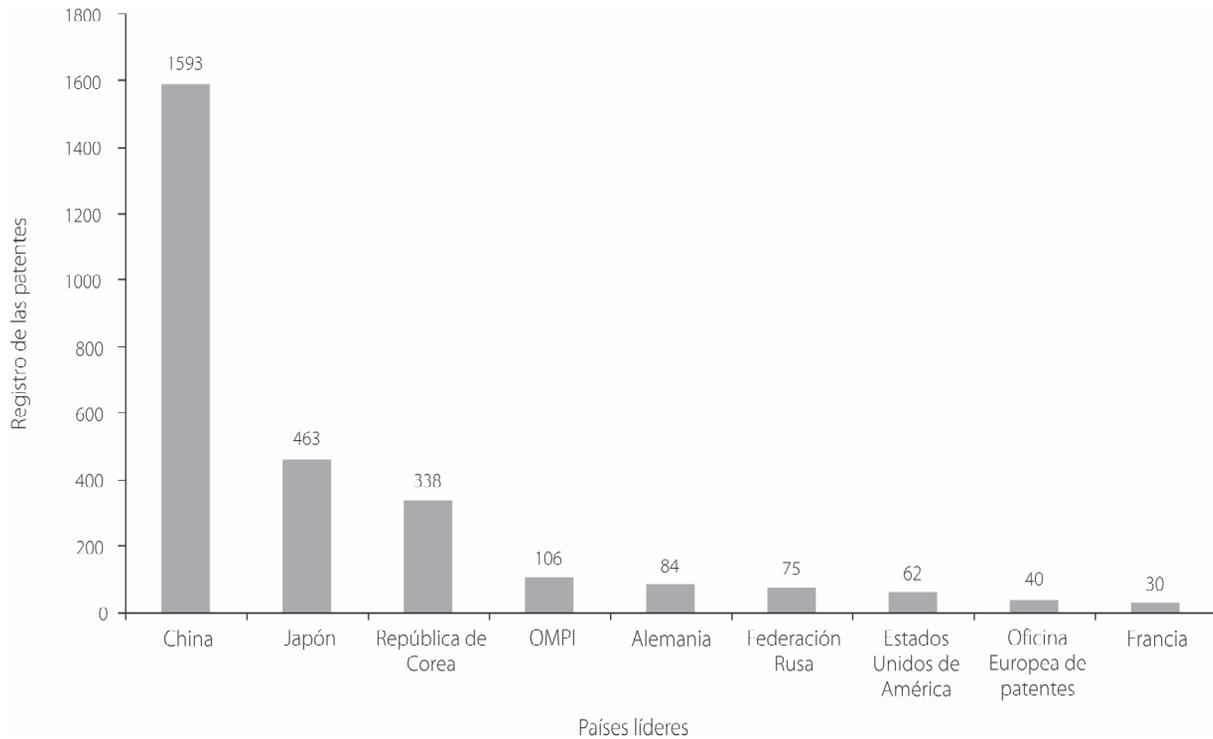


Figura 1. Países líderes.

Note: OMPI: Organización Mundial de la Propiedad Intelectual.

Fuente: Elaboración propia, 2013.

Las patentes incluyen información sobre las actividades y el rendimiento inventivo de las multinacionales. A través de las direcciones de los solicitantes y los inventores, es posible dibujar los patrones y medir la intensidad de la coinventión internacional. El análisis de los titulares líderes en la temática brindó como resultado que existe gran variedad de firmas implicadas en la producción de patentes relacionadas al tema (Figura 2). Principalmente son como se exponía más arriba coincidente con las tendencias mundiales, universidades y reconocidas transnacionales.

Los inventores son personas físicas, por lo general empleados del solicitante de la patente. En Estados Unidos, los inventores son los solicitantes. No obstante, también es posible que una misma persona sea el inventor y el solicitante (por ejemplo, el caso de los inventores independientes).

Los investigadores líderes a nivel mundial resultaron el señor Anil Jha de United States filter

Corporation con 50 solicitudes y *Nobuyuki Ashikaga de Toshiba* con 49 solicitudes. Un indicador muy importante del registro de patentes a nivel mundial es poder identificar dentro de un proceso tecnológico cuales son las variantes con mayor número de registros y su evolución en el tiempo.

En cuanto a las tecnologías de tratamiento del agua con más solicitudes utilizando los códigos de la CIP se detectan los siguientes (Tabla 1).

Estas son variantes tecnológicas que en el periodo 2006-2013, con más solicitudes presentan, por lo que afirmamos que son las tecnologías en pleno desarrollo o emergentes.

Se extraen otros indicadores importantes como son la evolución de cada una de estas variantes tecnológicas a nivel de subclase por año. Se observa una casi totalidad de registros en la subclase C02F (Tratamiento del agua y el agua residual) con presencia en todos los años estudiados, seguido de B01D (Métodos físico-químicos de separación) con presencia de más de

diez registros por año y llegando a 23 en el 2007. Se identifican dentro de los tratamientos biológicos los más destacados, siendo estos los tratamientos que emplean fangos activados (751) y los que usan microorganismos (624). Dentro de los de varias etapas se destacan los que al menos incluyen en una etapa un tratamiento biológico y los que incluyen uno químico (631). Se logra obtener un indicador muy importante pues combina temáticas, años de solicitud y cantidad de documentos, siendo un complemento de los indicadores anteriores. Se incluyen además los perfiles de firmas de las firmas líderes y de personalidad de los inventores líderes. Se incluyen los

resúmenes de la totalidad de las patentes y los textos completos de las más destacadas, tomando como parámetro para el grado de importancia las citas en posteriores documentos. Toda la información se presenta en una herramienta informatizada de presentación donde se incluyen los siguientes documentos: Breve descripción del servicio, la estrategia implementada, un informe basado en información no patente del tema de la búsqueda, el comportamiento de cada uno de los indicadores de patentes acompañado de la explicación del mismo, el resumen de las patentes, las patentes a texto completo de los más importantes, los perfiles de firmas y personalidades (Figura 3).

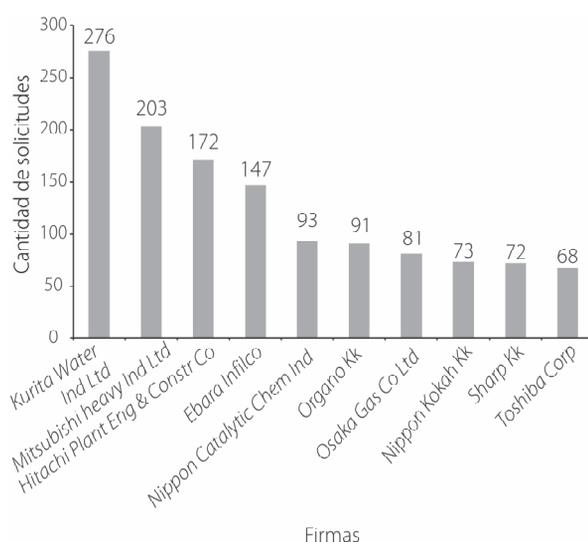


Figura 2. Firmas líderes.

Fuente: Elaboración propia, 2013.

Discusión

En la literatura se refieren numerosos artículos sobre el tratamiento del agua. Durante años se ha utilizado el cloro para la desinfección del agua con buen resultado. No obstante, el cloro reacciona con compuestos orgánicos presentes en las aguas y produce compuestos organoclorados y trihalometanos, los cuales tienen efectos mutagénicos y carcinogénicos. Además, se reportan microorganismos patógenos resistentes al cloro (Fernández García, 2013). Sin embargo, el empleo del ozono es una alternativa segura y eficaz de tratamiento de aguas, no se reportan los inconvenientes de la cloración. En este sentido se ha incrementado en las últimas décadas dado su alto poder oxidante y su elevado efecto germicida de amplio espectro contra bacterias, hongos, virus y quistes de parásitos, junto a la

Tabla 1. Tecnologías más registradas.

Código de la Clasificación Internacional de Patentes	Descripción	Cantidad de documentos encontrados
C02F1/44	Tratamiento del agua por diálisis ósmosis u ósmosis inversa	287
C02F1/52	Tratamiento del agua por floculación o precipitación de las impurezas en suspensión	248
C02F1/28	Tratamiento del agua por absorción o adsorción	248
C02F1/32	Tratamiento del agua por irradiación por luz ultravioleta	158
C02F1/50	Tratamiento del agua por adición o empleo de un germicida o por <i>tratamiento</i> oligodinámico	148
C02F1/78	Tratamiento del agua por oxidación por medio de ozono	145

Fuente: Elaboración propia, 2013.



Figura 3. Herramienta para visualizar la información.

Fuente: Worldwide... (2013).

producción de ozonizadores más compactos y eficientes. Se menciona la aplicación del método de oxidación por medio de ozono en procesos de potabilización de aguas, tratamientos de aguas residuales de diferente procedencia y para el tratamiento de agua de piscinas, también en aplicaciones a usos o procesos en los que se requiera un agua de elevada calidad como en la Industria farmacéutica y alimenticia. Este estudio corrobora la validez de esta tecnología y su factible uso detectado como una de las tecnologías emergentes en el presente estudio.

En las aguas que no se utilizarán en el consumo humano son otras las tecnologías más factibles de emplear, como pueden ser diferentes tratamientos biológicos. Se refiere la utilización de pasto vetiver, a pequeña escala el objetivo es estudiar el tratamiento complementario de aguas residuales en una planta de

residuales de fábrica de pepsi-cola (Scavo *et al.*, 2013). Este estudio no analiza otras tecnologías basadas en tratamientos biológicos también muy factibles a emplear. Otro estudio realizado en la Universidad de Stuttgart, Alemania; contando con unas noventa páginas, para el tratamiento de las aguas y su reuso en la capital del Perú específicamente del área metropolitana (McElhinney, 2012) se abordan las alternativas en cuanto a tecnologías para aplicar y no se menciona ni en las fuentes de información consultadas y en las tecnologías comparadas nada acerca de patentes que describen las tecnologías y que son la fuente primaria muchas veces.

Otro estudio (Radjenovic *et al.*, 2009) muestra los resultados del uso de tecnologías detectadas como la osmosis inversa *Reverse Osmosis* (RO) utilizada para la eliminación de residuales farmacéuticos. Este estudio tampoco menciona la información contenida en los

documentos de patentes como basamento teórico en la toma de decisiones a la hora de emplear un método u otro.

En otro documento (Reverse..., 2009) se plantea que cuando sólo se necesita un pequeño volumen de agua tratada, aproximadamente 0-10 galones por día, la osmosis inversa (RO) es normalmente el proceso más flexible y rentable de tratamiento disponibles para uso doméstico privado.

Por otro lado se plantea en cuanto al uso de tecnologías de oxidación con ozono (Ozone..., 2011) viable como desinfectante para matar microorganismos en el aire y el agua. Muchos spas y jacuzzis usan ozono para mantener el agua libre de algas y bacterias dañinas. El ozono también se utiliza con fines industriales y de fabricación, así como el blanqueo y agente desodorante. Las plantas de tratamiento de agua que fueron construidas en 1950 y 1960, respectivamente, se están actualizando para satisfacer las crecientes demandas de agua, a lo largo de las regulaciones de agua más estrictas federales y estatales, con el objetivo final de mejorar la calidad del agua. Históricamente, el cloro ha sido utilizado con eficacia en estas plantas de tratamiento de agua a desinfectar el agua potable. Sin embargo, con los avances en el tratamiento del agua, ha sido determinado que la desinfección con ozono es rentable y resultará en agua de mejor calidad.

Existe en la literatura numerosos estudios sobre tecnologías para el tratamiento de agua y agua residual, pero ninguno hace referencia a indicadores de patentes y documentos de patentes que describen las invenciones que solucionan las diferentes problemáticas tecnológicas que esta esfera presenta. Algunos estudios corroboran el grado de maduración de las tecnologías detectadas en nuestro estudio como tecnologías en auge, lo que indica la precisión del estudio de las patentes a la hora de localizar tecnologías para aplicaciones en transferencia de tecnología.

Indicadores de patentes como las firmas líderes nos sirven para analizar posibles socios comerciales, adquisición de licencias de patentes y otros tipos de negociaciones. Los demás indicadores ayudan a tener una visión global de la situación de desarrollo tecnológico en la materia que queramos investigar.

Es de vital importancia en las temáticas de información la manera que se presenta esta. A este efecto las tecnologías de la información y la comunicación juegan un papel primordial ofreciéndonos herramientas que a la distancia de un clic nos permite acceder a la información extractada y de alta elaboración como fungen los servicios de información que ofrecemos. El poder contar con herramientas de procesamiento, presentación y representación de la información logra hacer más atractivo el servicio, tener un producto más amigable y cómodo para el uso.

Conclusión

Se logra recuperar un número suficiente de soluciones tecnológicas contenidas en los documentos de patentes fundamentado por la estrategia de búsqueda trazada.

Se procesa la totalidad de los datos bibliográficos de las patentes utilizando para ello herramientas diseñadas en el centro de Información y Gestión tecnológica de Villa Clara y otras disponibles en la *Web*, extrayendo varios indicadores que permiten tomar decisiones en cuanto a tecnologías a utilizar en dependencia del uso posterior del agua y por supuesto el espacio disponible y los recursos financieros. También se identifican líderes mundiales en la generación de soluciones tecnológicas a la problemática en cuestión como es la firma japonesa *Kurita Water Industries Limited* y otras firmas e institutos de investigación con los que se puede negociar cualquier tipo de asociación. Se obtienen otros indicadores de actividad de países, inventores y otros de valor aplicable a actividades de centros de investigación y proyecto como patentes más citadas por otros autores lo que valoriza el documento como primario de una tecnología.

Toda la información recuperada es de gran utilidad para investigación y proyecto de plantas para tratamiento de agua y agua residual, cumpliendo así con los requisitos para ser la más valiosa fuente de información en la actividad de vigilancia tecnológica. Documentos de patentes relativos a equipos para decidir a la hora de proyectar equipamiento para tratamiento de agua y potabilización. Documentos relativos a procedimientos a la hora de decidir qué tipo de tratamiento emplear según su aplicación a tratamientos de aguas residuales de diferentes industrias.

Referencias

- Aleman, S. *Año internacional de la cooperación en la esfera del agua*. 2013. Disponible en: <<http://www.ecogestos.com/2013-ano-internacional-de-la-cooperacion-en-la-esfera-del-agua/>>. Acceso en: 15 enero 2013.
- European Patent Office. 2013. Available from: <http://worldwide.espacenet.com/advancedSearch?locale=en_EP>. Cited: Sept. 26, 2013.
- Fernández-Alba, A.R. et al. *Tratamientos avanzados de aguas residuales industriales*. Madrid: Universidad de Alcalá, 2006. (Vigilancia Tecnológica, v.2).
- Fernández García, L.A. *Tratamiento de aguas y aguas residuales con el empleo de ozono*. Centro de investigaciones del ozono. 2013. Disponible en: <<http://www.ozono.cubaweb.cu/acerca/tratamiento.htm>>. Acceso en: 22 oct. 2013.
- Intellectual Property Exchange. 2013. Available from: <www.ipexl.com/>. Cited: May 19, 2013.
- Keim, D.A. Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, v.8, n.1, p.18- 67, 2002.
- McElhinney, K.J. *Auditing water resources for application to water-sensitive urban design: A case study in the Lima (Perú) metropolitan area*. 2012. Available from: <http://www.lima-water.de/documents/kmcelhinney_tesis.pdf>. Cited: Dec. 17, 2013.
- Open source for innovation cartography (An). *Patent Lens*. 2013. Available from: <<http://www.lens.org/lens/>>. Cited: Dec. 12, 2013.
- Organización para la Cooperación y el Desarrollo Económicos. *Manual de estadísticas de patentes de la OCDE*. 2009. Disponible en: <http://www.oepm.es/export/sites/oepm/comun/documentos_relacionados/Memorias_de_Actividades_y_Estadisticas/EstadisticasOCDE/Estadisticas_OCDE.pdf>. Acceso en: 23 oct. 2013.
- Ozone disinfection for drinking water Miramar and Alvarado water treatment plants. 2011. Available from: <<http://www.sandiego.gov/cip/pdf/ozonefaq.pdf>>. Cited: Apr. 15, 2013.
- Patents. *CambridgeIP*. 2013. Available from: <http://www.boliven.com/patents/search?q=abstract%3A%28water+treatment%29&query_type=>>. Cited: May 20, 2013.
- Radjenovic, J. et al. Rejection of pharmaceuticals in nanofiltration and reverse osmosis membrane drinking water treatment. *Water Research*, v.42, n.14, p.3601-3610, 2008.
- Reverse osmosis treatment for drinking water. New Hampshire: Departmente Environmental Services, 2009. Available from: <<http://des.nh.gov/organization/commissioner/pip/factsheets/dwgb/documents/dwgb-2-11.pdf>>. Cited: May 7, 2013.
- Scavo, M.; Rodriguez, O.; Luque, O. *Estudio de un sistema de tratamiento de aguas residuales complementario, con pasto vetiver (vetiveria zizanioides L.), provenientes de una planta de producción de gaseosas, en villa de cura, estado Aragua*. 2013. Disponible en: <<http://www.vetiver.org/ICV4pdfs/BA17es.pdf>>. Acceso en: 19 dic. 2013.
- Trippe, A.J. Patinformatics: Tasks to tools. *World Patent Information*, v.25, n.3, p.211-221, 2003.
- Wikipedia. *Agua potable*. 2013. Disponible en: <http://es.wikipedia.org/wiki/Agua_potable>. Acceso en: 18 feb. 2013.
- Worldwide patent directory and citation statistics. *Patentmaps*. 2013. Available from: <<http://www.patentmaps.com/inventor/index.html>>. Cited: May 26, 2013.

Tecnologías constituidas, innovaciones en proceso y tecnologías introducidas en el mercado internacional: caso de estudio¹

Consolidated technologies, ongoing innovations and technologies introduced to the international market: A case study

Maidelyn DÍAZ-PÉREZ²

Raudel GIRÁLDEZ-REYES²

Dayron ARMAS-PEÑA²

Reinaldo Javier RODRÍGUEZ-FONT²

Elio Atenógenes VILLASEÑOR-GARCÍA³

Humberto Andrés CARRILLO-CALVET⁴

Resumen

La presente investigación tiene como objetivo conocer la trayectoria tecnológica del dominio caso de estudio Shale gas. Se propone como unidad de análisis y medición, la clasificación internacional de patentes de las invenciones relacionadas con el tema y recuperadas en tres bases de datos internacionales diferentes. Primero, se analizan las tecnologías constituidas consideradas como patentes concedidas en la oficina de patentes de los Estados Unidos de América *United States Patent and Trademark*, segundo las innovaciones en proceso identificadas como patentes solicitadas en la misma *United States Patent and Trademark*, y por último las patentes introducidas en el mercado internacional tramitadas mediante el Tratado de Cooperación de Patentes. La metodología utilizada aplicó indicadores de frecuencia y relacionales a la unidad objeto de estudio y obtuvo resultados que describen el comportamiento del conocimiento tecnológico del dominio. El análisis de la representación de los datos permitió identificar mediante el estudio de las clasificaciones las temáticas presentes en los procesos geofísicos de los aceites y gas de lutita conocidos como Shale gas, relacionados con la utilización de equipos de cómputo digital.

Palabras clave: Clasificación internacional de patentes. Indicadores de patentes. Patentes. Shale gas. Trayectorias tecnológicas.

Abstract

The aim of this research is to understand the technological trajectory of the case study of Shale gas production. As a unit of analysis and measurement, we used the international classification of patents for inventions related to the topic retrieved from three different

¹ Trabajo presentado en el VII Seminario Internacional sobre Estudios Cuantitativos y Cualitativos de la Ciencia y la Tecnología "Prof. Gilberto Sotolongo Aguilar" en XIII Congreso Internacional de Información - INFO' 2014. Habana, Cuba.

² Universidad de Pinar del Río, Grupo Gestión de Información, Conocimiento y Tecnologías, Dirección de Información Científico Técnica. Calle Martí Esquina 300, entre 27 de Noviembre y González Alcorta, 20100, Pinar del Río, Cuba. Correspondencia a nombre de/Correspondence to: M. DIAZ-PÉREZ. E-mail: <maidelyn@ict.upr.edu.cu>.

³ Centro de Investigación e Innovación en Tecnologías de la Información y Comunicación, Dirección Adjunta de Innovación y Conocimiento. Ciudad de México, México.

⁴ Universidad Nacional Autónoma de México, Facultad de Ciencias, Departamento de Matemática. Ciudad de México, México.

Recibido el día 26/5/2014 y aceptado para su publicación el 5/9/2014.

international databases. First the study analyzes consolidated technologies considered as granted patents, then the ongoing innovations identified as patent applications, and last the technologies introduced to the international market granted through the Patent Cooperation Treaty. According to the methodology, the frequency and relational indicators were applied to the unit under study. The results obtained describe the technological behavior of the domain. This technique allowed identifying the subject areas involved in the geophysical processes of oil and Shale-gas production related to the use of digital computing systems through the study of patent classifications.

Keywords: International patent classification. Patent indicators. Patents. Shale-gas production. Technological trajectories.

Introducción

Dentro de las normativas que establece la Organización Mundial de la Propiedad Intelectual (OMPI) se encuentra que toda invención debe acreditarse en una o varias Clasificaciones en el campo número 51 del documento de patente, las cuales identifican y describen a qué áreas del conocimiento técnico pertenece la invención. La Clasificación Internacional de Patentes (CIP) abarca un conjunto de campos de conocimiento incluidos dentro del ámbito de las patentes de invención. Al ser un sistema único e internacional logra una estandarización en las taxonomías del conocimiento tecnológico. Abarca no solo las grandes clases y categorías por disciplinas, sino hasta los niveles más profundos de complejidad de cada temática.

Los niveles jerárquicos de la última versión (octava edición) de la CIP con toda su estructura de símbolos, contiene: 8 secciones, 129 clases, 633 subclases, 7.066 grupos y 10.047 subgrupos. Esta amplia cobertura temática permite realizar análisis métricos considerando este campo del documento de patente; sin embargo no se utiliza con regularidad en los análisis métricos con la información.

En el análisis de literatura científica el análisis de las clasificaciones tampoco ha sido muy utilizado, se realizan más estudios de cocitación. Leydesdorff justifica este comportamiento al hecho que las clasificaciones o temáticas, son impuestas por las revistas, mientras las referencias citadas son elegidas por los autores. Aunque los indicadores relacionados con la clasificación son importantes, hasta el presente se han usado más los de cocitación. Muy a pesar de esas prácticas, el presente estudio rompe con ese esquema metodológico y presenta las amplias posibilidades que ofrece la clasificación para recuperar estructuras de conocimiento científico tecnológico (Díaz-Pérez, 2010).

Las clasificaciones pueden ser las unidades de análisis más confiables para explorar la información contenida en las bases de datos de patentes.

Esta investigación comparte el criterio que el análisis por clasificación temática permite mapear de forma más evidente el conocimiento científico que subyace en los documentos de patente (Leydesdorff, 1987; Tijssen, 1992; Leydesdorff & Heimeriks, 2001; Jaffe & Trajtenberg, 2002; Leydesdorff, 2004) así como permite conocer y representar la especialización temática por países, regiones, institutos, investigadores, etc. (Fall *et al.*, 2003; Van Zeebroeck *et al.*, 2006; Leydesdorff & Meyer, 2007; Cong & Han-Tong, 2010).

Y lo más importante, se considera que la CIP puede ser usada como unidad de análisis para explorar la información contenida en las patentes (Díaz-Pérez & Moya-Anegón, 2008), e incluso puede ser comparada con las categorías de materias que el *Institute for Scientific Information* (ISI) atribuye a las revistas según Leydesdorff. Este autor considera a las clases de la clasificación como equivalentes de las revistas, realiza la propuesta de analizar el equivalente correspondiente de una matriz de citación entre revista y revista (Leydesdorff, 2008).

Este estudio además de aplicar los indicadores tradicionales de patentes propone utilizar dentro de las unidades de análisis del estudio a la CIP. Se insiste en esta investigación en resaltar, que dentro de los estudios patentométricos el uso de la clasificación internacional es una de las mejores unidades de análisis y medida para diagnosticar la composición del estado de la técnica de un dominio tecnológico.

El caso de estudio que se utiliza para aplicar la propuesta está asociado con el dominio tecnológico Shale gas.

Por años, la extracción de reservas de petróleo y gas licuado han sido las fuentes de energía del mundo industrializado; no obstante, la creciente dependencia y

las, cada vez más escasas, reservas existentes han dado pie a la búsqueda de fuentes de energía no convencionales; países como Estados Unidos de América, Canadá, Europa y Australia, han emprendido esfuerzos por explotar las reservas de gas y aceite de lutita. Se estima que el gas de lutita, conjuntamente con otras fuentes de energía no convencionales logrará afianzar la independencia energética de estos países y reducir los elevados costos del crudo en el mercado internacional.

México no ha sido la excepción, a través del Boletín N° 26, emitido el 23 de marzo de 2011, Pemex manifestó haber obtenido su primera producción de gas de lutita pero con algunas limitaciones. Recientemente, el Congreso Mexicano aprobó un conjunto de modificaciones a la constitución denominado "Reforma Energética". Estas nuevas leyes prevén la explotación de este tipo de yacimientos a gran escala. Sin embargo, la producción del Shale gas tiene problemas técnicos y riesgos particulares. De los retos técnicos más importantes son la estimación no invasiva de las propiedades geológicas de los yacimientos y la predicción de su rentabilidad económica; ya que la construcción de los pozos implica una alta inversión económica, así como un alto riesgo al medio ambiente.

Para abordar el problema de caracterización no invasiva, se han propuesto algunas técnicas de modelación, tales como el mapeo geostadístico y transformaciones multiatributo para la estimación de la porosidad (Pramanik *et al.*, 2004). Otros enfoques incluyen el uso de algoritmos genéticos (Dorrington & Link, 2004) o redes neuronales (Helle *et al.*, 2001; Bhatt & Helle, 2002). Estas últimas han sido utilizadas exitosamente para estimar la porosidad y permeabilidad a partir de registros sísmicos. Recientemente, se presenta una aplicación de regresiones suaves para estimar la porosidad integrando atributos sísmicos con registros de pozos (Iturraran, 2011). Una de las ventajas de esta técnica es que además de estimar la porosidad se logran esclarecer relaciones no-lineales (si es que existen) entre las distintas variables que determinan la rentabilidad de los yacimientos, mediante el uso de simulaciones computacionales. Por esta razón, uno de los énfasis principales de esta investigación se hace sobre el desarrollo de tecnologías de la computación relacionadas; así como métodos de caracterización y simulación.

El objetivo de la presente investigación consiste en estudiar la trayectoria de este dominio tecnológico, a partir del análisis de la asignación de clasificadores de la CIP a patentes registradas en tres bases de datos diferentes *United States Patent and Trademark* (USPTO) concedidas, USPTO solicitadas, *PatentScope*. Se analizan las tecnologías constituidas (patentes concedidas), las innovaciones en proceso (patentes solicitadas) y las tecnologías introducidas en el mercado internacional tramitadas mediante el Tratado de Cooperación Internacional de patentes *Patent Cooperation Treaty* (PCT) del caso de estudio: Shale gas. La aplicación de indicadores de frecuencia y relacionales a la unidad objeto de estudio permite obtener resultados que describen el recorrido de este dominio, identificando mediante estas técnicas las temáticas presentes en los procesos geofísicos de aceite y gas de lutita (shale gas).

Métodos

La investigación considera que la unidad de análisis más representativa que será objeto específico de estudio en una medición; se refiere al qué o quién es objeto de interés en una investigación; es "el sobre qué o quiénes se van a recolectar datos". En este caso, lo que se analiza es un dominio tecnológico y la unidad de análisis propuesta es la CIP. Clasificación que cubre todos los ámbitos de la tecnología y detalla cada invención por niveles tan específicos que casi llega a describir los resultados.

Los indicadores utilizados son de frecuencia y relacionales. La unidad de análisis se cruza con otras variables (años, inventores, titulares, etc.) del documento de patente para obtener datos cruzados del dominio. Además se aplica el indicador patentométrico de coclasificación. Según lo establecido por la OMPI que consiste en:

Otra forma de establecer vínculos tecnológicos es a través de los códigos de clasificación. Esto se debe a que una misma patente puede ser clasificada bajo varios códigos de clasificación si se refiere a un producto o proceso que concierne a más de un campo de la tecnología. Los códigos de clasificación muestran los vínculos entre estos campos (Oficina Española de Patentes y Marcas, 1994, p.96).

En el caso del indicador de clasificaciones conjuntas será más o menos relevante en la medida en que ocurran en mayor o menor número de veces.

Como parte de la metodología se explica de forma independiente la fuente de datos y estrategia de búsqueda en cada base de datos de patentes utilizada (USPTO, Oficina de Patentes y Marcas de Estados Unidos y Organización Mundial de la Propiedad Intelectual (OMPI). Y al final se comenta de forma general la herramienta empleada para el procesamiento de la información y las técnicas de visualización utilizadas en los tres casos de análisis.

Primer análisis: *United States Patent and Trademark* concedidas

Fuente de datos: Oficina de Patentes y Marcas de Estados Unidos USPTO en específico se utiliza como fuente, la Base de Datos (BD) Patent Grant.

Estrategia de búsqueda: Interrogar la BD *Patent Grant de Estados Unidos* sobre los registros de patentes que tuvieran los términos "Shale gas" o "Shale oil", relacionados con: simulación, algoritmo, modelo, caracterización, reservorio y depósito.

Análisis: *United States Patent and Trademark* solicitadas

Fuente de datos: Oficina de Patentes y Marcas de Estados Unidos (USPTO) se utiliza en particular como fuente, la base de datos *Patent Application Full Text and Image Data Base* (solicitadas).

Estrategia de búsqueda: Interrogar la BD *Patent Application Full Text and Image Data Base* de Estados Unidos sobre los registros de patentes que contengan los términos "Shale gas" o "Shale Oil", relacionados con: simulación, algoritmo, modelo, caracterización, reservorio y depósito.

PatentScope

Fuente de datos: fue consultada la base de datos de patentes tramitadas por vía PCT, gestionada por la OMPI mediante la herramienta informática PatentScope®.

Estrategia de búsqueda: Interrogar la BD de PatentScope® sobre los registros de patentes que

contengan los términos "Shale gas" o "Shale oil", relacionados con: computación, algoritmo, simulación.

Procesamiento de los datos

Se utiliza el *software proINTEC* (Giráldez *et al.* 2008) para la descarga, normalización, procesamiento, análisis y visualización de los datos, con confiabilidad probada en otros estudios de caso (Díaz-Pérez, *et al.*, 2007; Díaz-Pérez, *et al.*, 2008; Díaz-Pérez, *et al.*, 2009; Díaz-Pérez, *et al.*, 2010). Para el análisis y visualización de la información contenida en las patentes PCT se usó el servicio gratuito ofrecido por la OMPI, mediante "PatentScope".

Técnicas de Visualización

Se utiliza como técnica de visualización las redes presentadas en forma de nodos o vértices, donde la estructura semántica se encuentra definida por los enlaces que conectan dichos nodos. La visualización mediante redes hace posible representar las relaciones en forma de red, cruzando variables para encontrar comportamientos o patrones dentro del contenido de las patentes. Para una mejor representación de la información en el análisis e interpretación de los mapas tecnológicos se utiliza el algoritmo *Pathfinder* que permite mostrar sólo los enlaces más relevantes de las redes mapeadas. Se usa además como técnica de poda de enlaces la delimitación numérica de la intensidad de las relaciones entre aquellos nodos que su relación es débil. Y para el posicionamiento de los nodos en la visualización del dominio se emplea el algoritmo de representación espacial *Spring Embeded*; por último se usa el programa *Netdraw* para manipular los gráficos en las representaciones visuales.

Resultados

Base de Datos *United States Patent and Trademark* concedidas

El dominio tecnológico Shale gas está compuesto en la USPTO por un total de 26 patentes concedidas en un periodo de tiempo de 36 años, comprendido entre 1976 y julio del 2012. La década del ochenta fue la

primera en exhibir un incremento del número de invenciones concedidas sobre el tema, y hasta el presente no ha sido superado.

Clasificación Internacional de Patentes

A partir de un análisis de frecuencia del campo clasificaciones, se obtuvo el dato que las patentes del dominio Shale gas abordan 5 secciones de la técnica. El mayor número de patentes se clasifican en la sección E "Construcciones Fijas". Sin embargo, las patentes con esta clasificación son en su gran mayoría de los años setenta y ochenta del pasado siglo, no representan novedad para este estudio. Sólo existe una patente con esta clasificación concedida en este siglo, registrada en el año 2001.

Las patentes de las clasificaciones F (Mecánica, etc.) y B (Técnicas Industriales Diversas) fueron también concedidas en los años ochenta del pasado siglo.

Las patentes concedidas después del año 2001, se suponen sean las más novedosas, pertenecen a las áreas Física y Química. Puede significar que en la actualidad las innovaciones científico tecnológicas sobre Shale gas relacionadas con tecnologías de cómputo digital estén vinculadas con las secciones G (Física) y C (Química) de la CIP.

El análisis por el *segundo nivel jerárquico de la CIP* reveló que las invenciones sobre Shale gas se ubican en 14 clases de la CIP. Sin embargo, por la novedad de la temática que abarca este estudio las más importantes son las trabajadas en el presente siglo.

Las clases temáticas del dominio concedidas después del año dos mil, son:

- C10: dentro de la Química abarca los conocimientos relacionados con "Industrias del petróleo, gas o coque. gas de síntesis que contiene monóxido de carbono. combustibles. lubricantes. turba".

- C1: dentro de la Química utiliza los conocimientos relacionados con "Química Inorgánica".

- C12: dentro de la Química estudia los conocimientos relacionados con "Bioquímica. cerveza. bebidas alcohólicas. vino. vinagre. microbiología. enzimología. técnicas de mutación o de genética".

- E21: dentro de la sección de Construcciones Fijas abarca los conocimientos relacionados con "Perforación del suelo o de la roca. explotación minera".

- G1: dentro de la sección Física contempla los conocimientos relacionados con "Metrología. ensayos".

- G6: dentro de la sección Física contiene los conocimientos relacionados con "Cómputo. cálculo. conteo".

Las clases C1, C10 y G6 fueron identificadas en las patentes concedidas sobre Shale gas por la USPTO en el año 2012.

El análisis por el *último nivel jerárquico de la CIP* mostró que existe un amplio número de subgrupos relacionados con las invenciones de Shale gas.

A continuación se presentan los subgrupos temáticos de donde pertenecen las patentes concedidas a partir del año dos mil:

C10G0009: Dentro de la Química abordan temas relacionados con "Cracking térmico no catalítico, en ausencia de hidrógeno, de aceites de hidrocarburos".

C10G0009/320000: Dentro de la Química abordan temas relacionados con "Cracking térmico no catalítico, en ausencia de hidrógeno, de aceites de hidrocarburos según la técnica de lecho fluidizado".

C10G0011/180000: Dentro de la Química abordan temas relacionados con "Cracking catalítico, en ausencia de hidrógeno, de aceites de hidrocarburos" según la técnica fluidizada.

C10G0051/060000: Dentro de la Química abordan temas relacionados con "Tratamiento de los aceites de hidrocarburos, en ausencia de hidrógeno, único por varios procesos de cracking, por varias etapas en paralelo".

C10G0073/020000: Dentro de la Química abordan temas relacionados con "Obtención de ceras de petróleo a partir de aceites de hidrocarburos. Desparafinado de aceites de hidrocarburos".

C12P0003: Dentro de la Química abordan temas relacionados con "Procesos de fermentación o procesos que utilizan enzimas para la síntesis de un compuesto químico dado o de una composición dada, o para la separación de isómeros ópticos a partir de una mezcla racémica", en específico preparación de elementos o compuestos inorgánicos excepto anhídrido carbónico.

C12P0005/020000: Dentro de la Química abordan temas relacionados con "Preparación de hidrocarburos. Acíclicos".

C1B0003/320000: Dentro de la Química abordan temas relacionados con "Hidrógeno. Mezclas gaseosas que contienen hidrógeno. Separación del hidrógeno a partir de mezclas que lo contienen" en específico en Purificación del hidrógeno por reacción de compuestos orgánicos gaseosos o líquidos con agentes gasificantes, p. ej. agua, dióxido de carbono, aire.

E21B0025/080000: Dentro del área temática de Construcciones Fijas abordan temas relacionados con "Aparatos para obtener o retirar las muestras intactas, p. ej. tubos de muestras, extractores de muestras" "Revestimiento, congelación, consolidación de muestras. Recuperación de muestras no contaminadas o de muestras a la presión de formación".

E21B0049: Dentro del área temática de Construcciones Fijas abordan temas relacionados con "Ensayos para determinar la naturaleza de las paredes de los orificios de perforación. Ensayo de las capas. Procedimientos o aparatos para obtener muestras de terreno o de los fluidos del pozo, adaptados a perforaciones del suelo o de pozos".

E21B0049/020000: Dentro del área temática de Construcciones Fijas abordan temas relacionados con "Ensayos para determinar la naturaleza de las paredes de los orificios de perforación. Ensayo de las capas. Procedimientos o aparatos para obtener muestras de terreno o de los fluidos del pozo, en principio adaptados a perforaciones del suelo o de pozos", en especial "Mediante muestras tomadas mecánicamente del terreno".

G1N0015/080000: Dentro del área temática de la Física se abordan temas relacionados con "Investigación de la permeabilidad, del volumen de los poros o del área superficial de los materiales porosos".

G1N0033/260000: Dentro del área temática de la Física se abordan temas relacionados con "Investigación o análisis de materiales por métodos específicos no cubiertos por los grupos", "aceites. líquidos viscosos. pinturas. tintas".

G1N0033/280000: Dentro del área temática de la Física se abordan temas relacionados con "Investigación

o análisis de materiales por métodos específicos no cubiertos por los grupos. Aceites".

G1V0003/060000: Dentro del área temática de la Física se abordan temas relacionados con "Prospección o detección eléctrica o magnética. Medida de las características del campo magnético terrestre, p. ej. de la declinación, de la desviación que utiliza corriente alterna".

G6F0019: Dentro del área temática de la Física se abordan temas relacionados con "tratamiento de datos digitales electricos", en concreto "Equipos o métodos de proceso de datos o de cómputo digital", adaptados a aplicaciones específicas.

Relaciones temáticas: clasificaciones conjuntas

El dominio Shale gas tiene relaciones entre diferentes secciones del conocimiento tecnológico. Las relaciones interdisciplinarias más fuertes ocurren entre la sección de Física (G) con la sección de Construcciones Fijas (E); y entre la sección de Química (C) con la sección de Técnicas industriales diversas (B). Existe también un núcleo de relaciones interdisciplinarias entre las secciones de Química (C), Mecánica (F) y Construcciones Fijas (E); así como esta última con el sector de la Física (Figura 1).

En el análisis de las clasificaciones conjuntas por el segundo nivel jerárquico de la CIP se destaca la

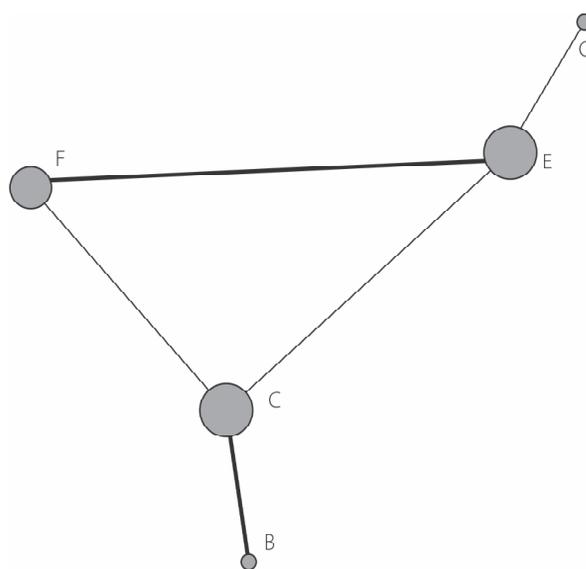


Figura 1. Clasificaciones conjuntas (Sección, Clasificación Internacional de Patentes).

Fuente: Software proINTEC 2013.

intensidad de las relaciones que existen entre las clases *F2* (motores de combustión, plantas motrices de gases calientes o de productos de combustión), la clase *E21* (perforación del suelo o de la roca, explotación minera) y la *F23* (aparatos de combustión, procesos de combustión).

Otra relación conjunta de interés para esta investigación se establece entre las clases *G6* orientada a cómputo, cálculo, conteo, etc., con la clase *G1* que aborda la metrología y a su vez, esta última relacionada con la *E21* que abarca los conocimientos técnicos sobre perforación de suelos o rocas, explotación minera, etc. Esto constata que existen patentes en este dominio que están reivindicando invenciones sobre tecnologías orientadas a la metrología y a la explotación de suelos y rocas.

El análisis hasta el último nivel jerárquico de clasificación representa las principales clasificaciones conjuntas del dominio (Figura 2). Se visualizan los diferentes grupos temáticos en Shale gas, incluso se delimitan los diferentes nodos que existen incluso dentro de una misma sección del conocimiento tecnológico. En este caso, las relaciones más intensas se manifiestan entre

la sección E de "construcciones fijas" en la temática de "perforación del suelo o de la roca, explotación minera", del área de "procedimientos o dispositivos para la extracción de petróleo, gas, agua o materiales solubles o fundibles o de una suspensión de materias minerales a partir de pozos, aquí las relaciones se establecen entre las patentes que clasifican en la temática relacionada con los "procedimientos de fracturación" (*E21B0043/247000*) y las patentes que clasifican en "procedimientos de recuperación optimizada para la extracción de hidrocarburos" (*E21B0043/160000*). Y esta última, con las patentes que clasifican en la temática "explosivos" (*E21B0043/248000*). Entre estos 3 temas existen el mayor número de clasificaciones conjuntas en el dominio Shale gas.

Las relaciones interdisciplinarias de mayor interés para esta investigación son las relaciones conjuntas entre las clasificaciones: *G1N0033-280000/G1N0033-260000/G6F0019*. Son importantes porque revelan que existen patentes en este dominio que han trabajado en tecnologías relacionadas con investigaciones o análisis de materiales por determinación de sus propiedades químicas o físicas, utilizando el tratamiento de datos digitales eléctricos, a partir de equipos o métodos de proceso de datos o de cómputo digital.

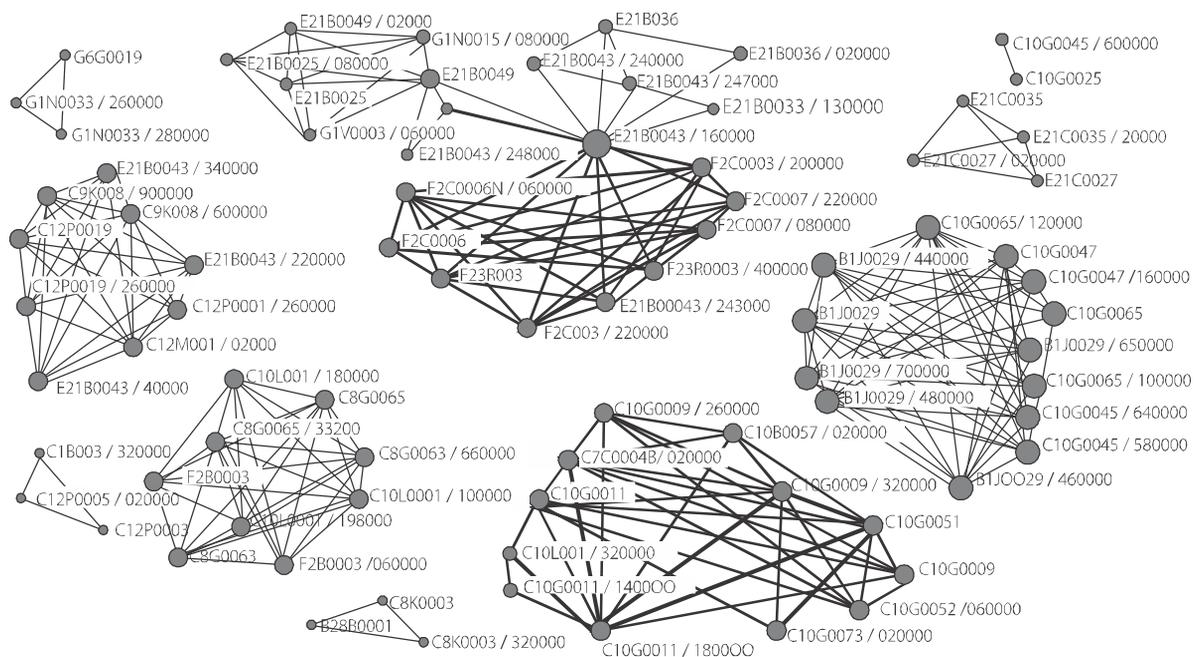


Figura 2. Clasificaciones conjuntas (Subgrupo, Clasificación Internacional de Patentes).

Fuente: Software proINTEC 2013.

Área temática de los inventores:

La sección de Química (C) y Construcciones Fijas (E) tiene el mayor número de investigadores asociados, seguida por la sección de Mecánica (F), Física (G) y Técnicas Industriales Diversas (B).

A partir del interés que tiene este estudio en el área de Física, a continuación se listan los investigadores que trabajan en esta área del conocimiento en el dominio Shale gas:

1. McNeil (III); Robert Irving
2. Nicholson; James William
3. O'Meara (Jr.); Daniel J.
4. Skelt; Christopher
5. Zeilinger; Sabine Claudia
6. dos Santos; Helio Mauricio Ribeiro

Área temática de los titulares:

El mayor número de titulares de este dominio trabajan en investigaciones tecnológicas relacionadas

con Construcciones Fijas (E), seguidos por el área de Química, así como los que investigan en ambas secciones del conocimiento tecnológico. En la sección G, F y B sólo le han concedido patentes a un reducido número de titulares en el periodo estudiado (Figura 3).

Existen sólo 2 titulares trabajando en el área de Física dentro del dominio, estos son: *Chevron Corporation* y *Shell Oil Company*. Y los titulares que ostentan mayor actualidad (patentes concedidas en 2012) en el dominio son: *Chevron U.S.A. Inc.* Con una patente en el año 2012 relacionada con la sección G (CIP); y *LUCA Technologies, Inc.*

Colaboración Científico Tecnológica

Este análisis revela que existen relaciones de colaboración entre los inventores que trabajan en las investigaciones patentables sobre Shale gas (Figura 4). La investigación resalta dentro del dominio el nodo de investigadores relacionados con el área de la Física. En este caso, los investigadores Robert Irving McNeil (III),

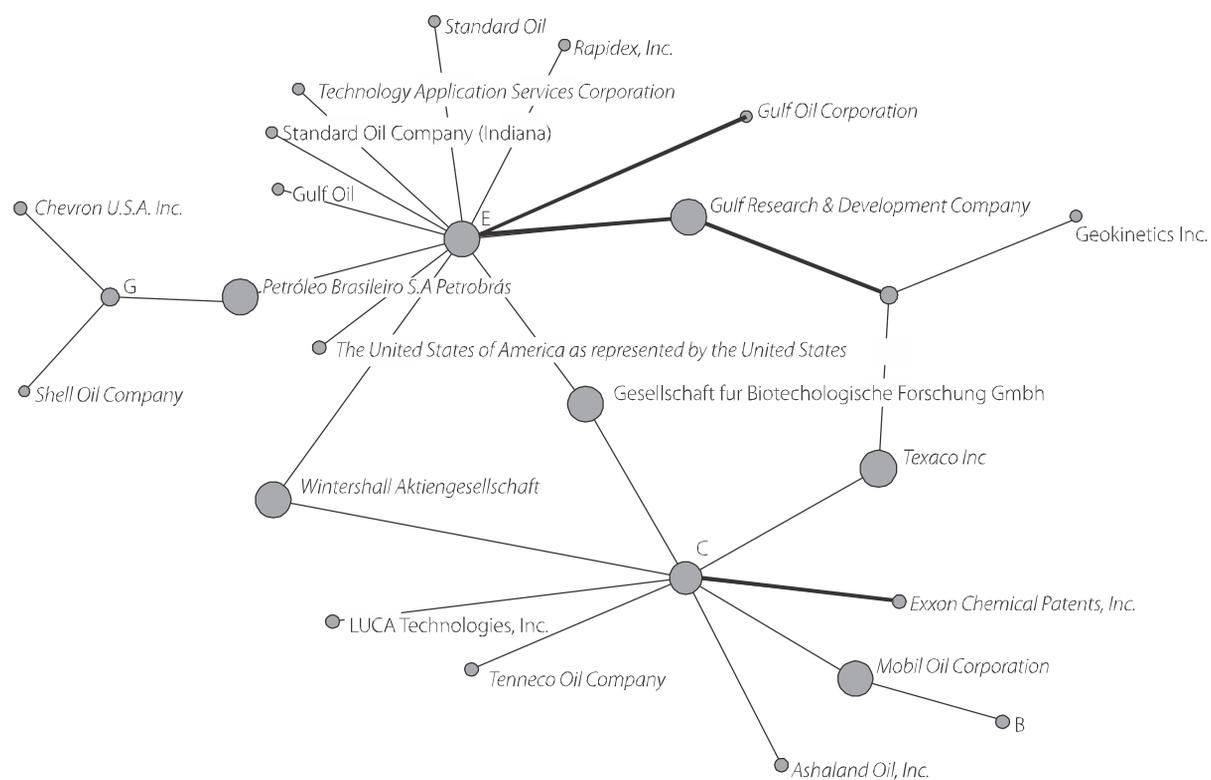


Figura 3. Titulares por Sección, Clasificación Internacional de Patentes.

Fuente: Software proINTEC 2013.

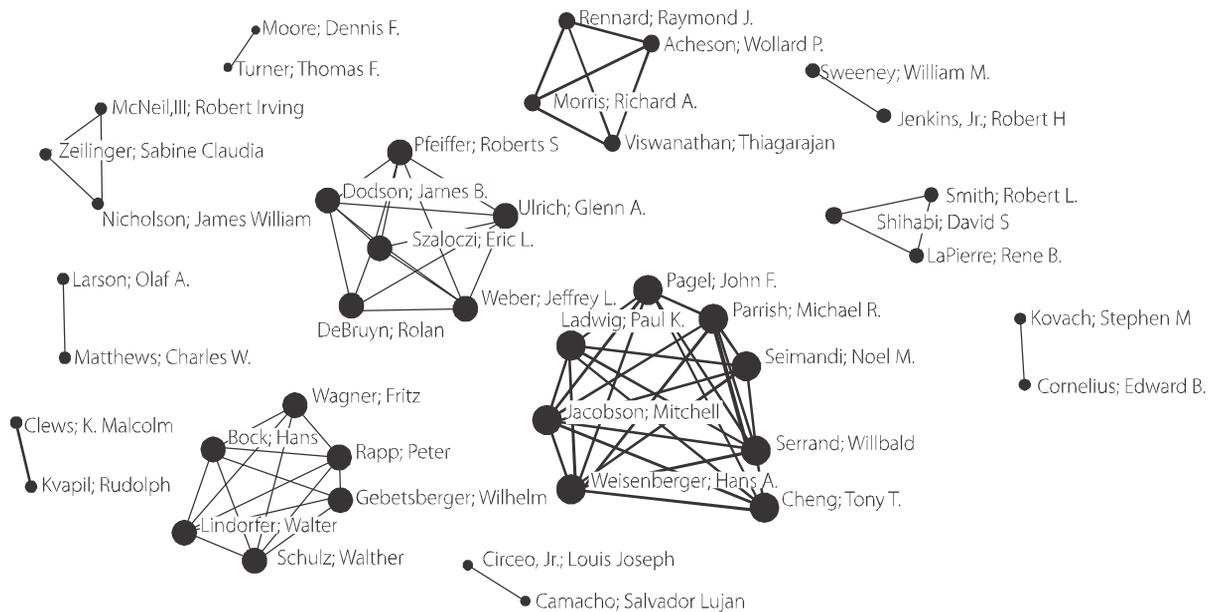


Figura 4. Inventores conjuntos.

Fuente: Software proINTEC 2013.

James William Nicholson y Sabine Claudia Zeilinger trabajan dentro del área técnica de Física en el "Tratamiento de datos digitales eléctricos", en específico "Equipos o métodos de proceso de datos o de cómputo digital" asociados a la búsqueda de Shale gas.

United States Patent and Trademark solicitadas

El dominio tecnológico sobre Shale gas en la BD de patentes solicitadas de Estados Unidos tiene un total de 51 patentes, entre el año 2001 y julio del 2012. El promedio de solicitudes al año es de unos 6 registros, se supera exponencialmente esta cifra en el año 2003 (28 solicitudes). Y se vuelve a observar un incremento de solicitudes a partir del 2010, crecimiento que se mantiene hasta el año 2012.

Clasificación internacional de patentes

Las patentes solicitadas ante la USPTO en este dominio trabajan en 5 secciones de la CIP. La clasificación de la sección E sobre *construcciones fijas* tiene el mayor número de patentes asociadas, seguidas por solicitudes

en: Física (13), Química (4), Técnicas industriales diversas (B) y Electricidad (2).

Los análisis por los diferentes niveles jerárquicos de la CIP destacan que:

- las solicitudes de patentes sólo trabajan en 9 clases tecnológicas. Aunque sus propuestas de invención abarquen 5 áreas tecnológicas de las 8 que existen en la CIP, sus investigaciones sólo se profundizan en una parte de estos sectores del conocimiento.

- las solicitudes trabajan en 11 subclases temáticas, donde se mantiene la sección G como la más investigada en sus diferentes sectores seguida respectivamente por las secciones: C, E, B y H.

En el análisis por *grupo* se amplían más las áreas de la tecnología que investigan cada sector tecnológico, y se destaca la incidencia científico-tecnológico que tienen los siguientes grupos:

- E21B0043: Procedimientos o dispositivos para la extracción de petróleo, gas, agua o materiales solubles o fundibles o de una suspensión de materias minerales a partir de pozos.

- G6G0007: Dispositivos en los que la operación de cálculo es efectuada haciendo variar valores eléctricos o magnéticos.

- E21B0036: Disposiciones para el calentamiento, el enfriamiento, o el aislamiento, en los orificios de perforación o en los pozos, p. ej. para utilizarse en zonas heladas.

- G6F0019: Equipos o métodos de proceso de datos o de cómputo digital, especialmente adaptados a aplicaciones específicas.

- E21B0047: Prospecciones en los orificios de perforación o en los pozos.

Por último, el análisis por subgrupos reveló que los principales grupos que derivan en subgrupos, son:

- E21B0043/200000: trabaja dentro de los procedimientos, que son mediante desplazamientos por agua.

- G6G0007/400000: Dispositivos en los que la operación de cálculo es efectuada haciendo variar valores eléctricos o magnéticos, de ecuaciones diferenciales parciales.

- E21B0043/300000: Procedimientos o dispositivos para la extracción de petróleo, gas, agua o materiales solubles o fundibles o de una suspensión de materias minerales a partir de pozos. Disposición particular de

pozos, p. ej. disposición óptima de espaciamiento de pozos.

- G6G0007/500000: Dispositivos en los que la operación de cálculo es efectuada haciendo variar valores eléctricos o magnéticos. para redes de distribución, p. ej. para fluidos.

- G1V0001/300000: Sismología. Prospección o detección sísmica o acústica. Análisis.

Se obtuvo además, que los titulares que solicitan invenciones en la sección de Física se corresponden con los titulares que tienen mayor productividad del dominio. Significa que las principales compañías del mundo liderando innovaciones sobre Shale gas investigan en esta área del conocimiento tecnológico.

Colaboración - En el dominio patentes solicitadas de la USPTO no se manifiestan muchas relaciones de colaboración entre titulares. Por ejemplo, sólo existen relaciones entre 2 signatarios: *China Petroleum & Chemical Corporation* y *Research Institute of Petroleum Processing Sinopec*.

Tampoco existen muchos grupos de investigación que colaboran entre sí en este dominio.

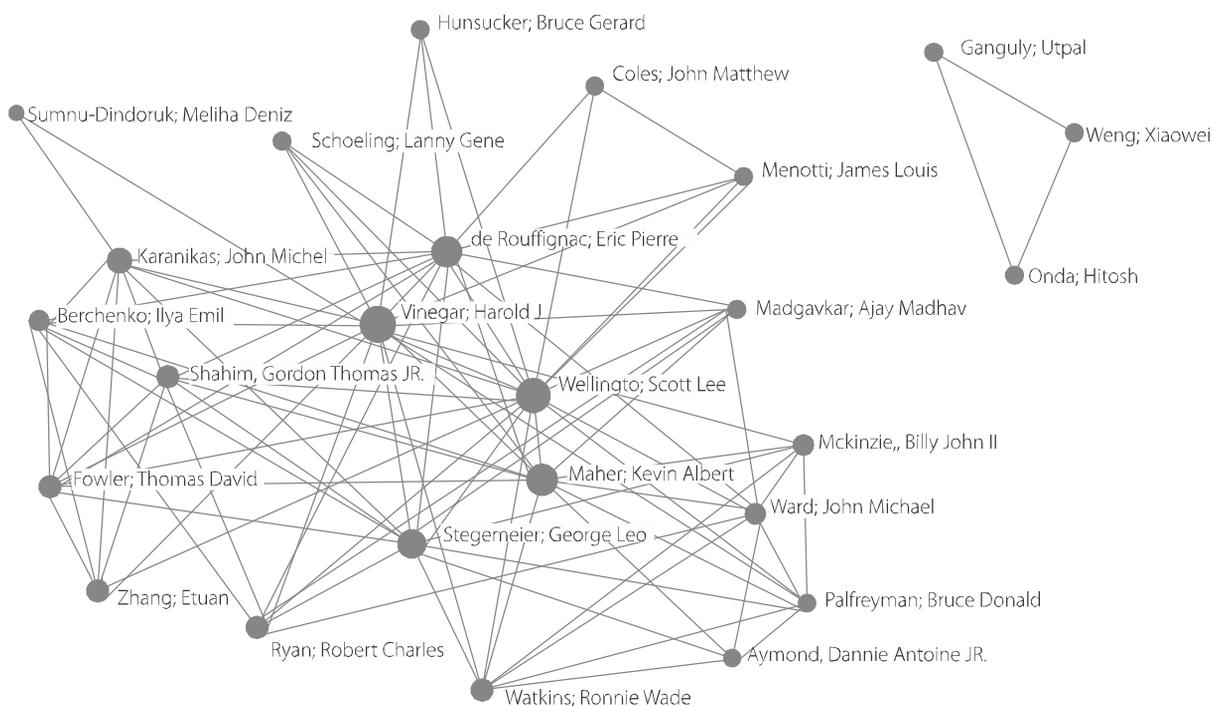


Figura 5. Colaboración de inventores.

Fuente: Software proINTEC 2013.

Solo se observa un amplio equipo de inventores pertenecientes a Shell Oil Company que colaboran entre sí para desarrollar las tecnologías que este titular solicita (Figura 5). Fuera de este amplio grupo, existen sólo 3 tecnólogos en colaboración, y como dato interesante estos inventores han solicitado protección en el año 2012.

PatentScope

En la base de datos *PatentScope* solo existen 11 patentes relacionadas con Shale gas. Cifra que indica un bajo grupo de invenciones tramitadas mediante el PCT en este dominio temático.

Dentro de la serie temporal que se analiza solo en 7 años se formalizaron patentes sobre Shale gas mediante PCT. Se observa que a partir del 2007 se incrementó la introducción de sus tecnologías en el mercado internacional, pero sufre una recaída en el 2010.

Clasificación Internacional de Patentes

Las 11 patentes de este estudio están relacionadas con 4 secciones de la CIP. El mayor número de patentes se relaciona con la temática (E21B) "Earth or rock drilling; obtaining oil, gas, water, soluble or meltable materials or a..." El resto de las clasificaciones pertenecen a la siguiente área de la técnica: C01B - Non-metallic elements; compounds thereof (1 patente), E21C - Mining or quarrying (1 patente), y G06G - Analogue computers (1 patente)

Conclusión

Se analizaron las tres bases de datos, las tecnologías constituidas mediante las patentes concedidas, las innovaciones en proceso a través de las invenciones solicitadas y las tecnologías introducidas en el mercado internacional mediante la vía PCT. Se identificó la trayectoria tecnológica del dominio Shale gas a partir del estudio de la unidad de análisis e indicadores propuestos. Los resultados de este análisis

cubren distintos aspectos tales como: relaciones temáticas, áreas temáticas de los inventores, colaboración científico-tecnológica y colaboración entre titulares.

Respecto a la colaboración se observa muy pocas relaciones de colaboración entre titulares. También se aprecia que existen muy pocos grupos de investigación que colaboran entre sí. Solo se identifica un amplio equipo de inventores pertenecientes a Shell Oil Company.

Con relación a la introducción de estas tecnologías en el mercado internacional se observa que ésta se ha dado principalmente en los últimos 7 años, en los cuales se formalizaron patentes sobre Shale gas mediante PCT.

El principal resultado de este análisis es que constató que existen patentes dentro de este dominio que están reivindicando invenciones sobre tecnologías orientadas a la metrología y a la explotación de suelos y rocas. Se pudo apreciar que se están desarrollando tecnologías relacionadas con análisis de materiales por determinación de sus propiedades químicas o físicas, utilizando el tratamiento de datos digitales; a partir de equipos o métodos de procesos de cómputo; son las grandes compañías como: Chevron Corporation y Shell Oil Company, las que están liderando estas innovaciones. Es decir, mediante la metodología propuesta se logran identificar las temáticas presentes en los procesos geofísicos de aceite y gas de lutita (Shale gas) relacionados con la utilización de equipos de cómputo digital, satisfaciendo el principal interés del estudio.

Agradecimientos

Los autores Elio A. Villaseñor y Humberto Carrillo agradecen al proyecto Prototipo de Plataforma interactiva para la Visualización y Simulación de Procesos Geofísicos de Gas y Aceite de lutita (Shale) con clave CONACYT/PROINOVA 184601 y a la empresa Servicios y Suministros en Informática S.A. de Capital Variable <www.plenumsoft.com.mx> por las facilidades brindadas para la realización de este trabajo.

Referencias

Bhatt, A.; Helle, H.B. Committee neural networks for porosity and permeability prediction from well logs. *Geophysical Prospecting*, v.50, n.645, p.660, 2002.

Cong, H.; Han-Tong, L. Pattern-oriented associative rule-based patent classification. *Expert Systems with Applications*, v.37, n.3, p.2395-2404, 2010.

Díaz-Pérez, M. *Visualización del análisis del dominio tecnológico de Cuba: 1997-2008*. 2010. Tesis (Doctoral en Documentación) - Departamento de Biblioteconomía y Documentación, Universidad de Granada, Granada, 2010.

Díaz-Pérez, M.; Giraldez Reyes, R.; Armas Peñas, D. Principales resultados de innovación tecnológica de Cuba en Estados Unidos: una visión desde las patentes. *Ciencias de la Información*, v.39, n.2, p.37-46, 2008.

Díaz-Pérez M., et al. Análisis patentométrico de la información desde la perspectiva de género. *Acimed*, v.20, n.2, 2009. Disponible en: <http://scielo.sld.cu/scielo.php?pid=S1024-94352009000800002&script=sci_arttext>. Acceso en: 20 sep. 2013.

Díaz-Pérez, M.; Guzmán M.V.; Orea U. Estudio patentométrico de un proyecto de investigación. *Ciencias de la Información*, v.38, n.1-2, p.57-66, 2007.

Díaz-Pérez, M.; Moya-Anegón, F. El Análisis de patentes como estrategia para la toma de decisiones innovadoras. *El Profesional de la Información*, v.17, n.3, 2008. Disponible en: <<http://www.elprofesionaldeinformacion.com/contenidos/2008/mayo/05.html>>. Acceso en: 2 oct. 2013.

Díaz-Pérez, M.; Rivero, S.; Moya-Anegón, F. Producción tecnológica latinoamericana con mayor visibilidad internacional: 1996-2007: un estudio de caso: Brasil. *Revista Española de Documentación Científica*, v.33, n.1, p.34-62, 2010.

Dorrington, K.P.; Link, C.A. Genetic-algorithm/neural-network approach to seismic attribute selection for well-log prediction. *Geophysics*, v.69, n.1, p.212-221, 2004.

Fall, C.J., et al. Automated categorization in the international patent classification. *ACM SIGIR Forum*, v.37, n.1, p.10-25, 2003.

Giráldez, R.; Díaz-Pérez, M.; Armas, D. PROInTec: un software para el tratamiento inteligente de datos sobre patentes. *Acimed*, v.17, n.5, p.1-9, 2008. Disponible en: <http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352008000500006>. Acceso en: 17 sep. 2013.

Helle, H.B.; Bhatt, A.; Ursin, B. Porosity and permeability prediction from wireline logs using artificial neural networks:

a north sea case study. *Geophysical Prospecting*, v.49, n.4, p.431-444, 2001.

Iturriaran, U. Smooth regression to estimate effective porosity using seismic attributes. *Journal of Applied Geophysics*, v.76, p.1-12, 2012.

Jaffe, A.B.; Trajtenberg, M. *Patents, citations & innovations: A window on the knowledge economy*. Cambridge, MA: MIT Press, 2002.

Leydesdorff, L. Various methods for the mapping of science. *Scientometrics*, v.1, n.5-6, p.291-320, 1987.

Leydesdorff, L. The university-industry knowledge relationship: Analyzing patents and the science base of technologies. *Journal of the American Society for Information Science & Technology*, v.55, n.11, p.991-1001, 2004.

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

Leydesdorff, I.; Heimeriks, G. The self organization of the European Information Society: The case of "Biotechnology". *Journal of the American Society for information Science and Technology*, v.52, n.14, p.1262-1274, 2001.

Leydesdorff, L.; Meyer, M. The scientometrics of a triple helix of university-industry-government relations (Introduction to the topical issue). *Scientometrics*, v.70, n.2, p.207-222, 2007.

Oficina Española de Patentes y Marcas. *Manual de estadísticas de patentes de la OCDE*. Madrid: OEPM, 1994.

Pramanik, A.G., et al. Estimation of effective porosity using geostatistics and multiattribute transforms: A case study. *Geophysics*, v.69, n.2, p.352-372, 2004.

Tijssen, R.J.W. *Cartography of science: Scientometric mapping with multidimensional scaling methods*. Leiden: DSWO Press, 1992.

Van Zeebroeck, et al. Issues in measuring the degree of technological specialization with patent data. *Scientometrics*, v.66, n.3, p.481-492, 2006.

Índice de autores

A

ABAD-GARCÍA, Maria Francisca	133
AGUIRRE-LIGÜERA, Natalia	155
ALCÁNTARA, Frank Coelho de	27
ANTUNES, Michele Nacif	9
APPEL, Andre Luiz	27
ARMAS-PEÑA, Dayron	349
AQUINO, Mirian de Albuquerque	203

B

BALLAGAS-FLORES, Bárbara de los Milagros	315
BETTONI, Eduardo Michellotti	27
BEUREN, Fernanda Hansch	167
BRÄSCHER, Marisa	191
BRETTAS, Aline Pinheiro	51
BUFREM, Leilah Santiago	295

C

CALDERA-SERRANO, Jorge	125
CAMPOS, Lucila Maria de Souza	77
CAREGNATO, Sônia Elisa	177
CARRILLO-CALVET, Humberto Andrés	327, 349
CAVALCANTE, Raphael da Silva	191
CERETTA, María Gladys	155
CHINCHILLA-RODRÍGUEZ, Zaida	281
CLARO-PÉREZ, Mirelys	339
COLLAZO-ALFONSO, Lisbey	339
COLLAZO-REYES, Francisco	269

D

DÍAZ-PÉREZ, Maidelyn	327, 349
DIGIAMPIETRI, Luciano Antonio	239

E

ENSSLIN, Leonardo	77
ENSSLIN, Sandra Rolim	77

F

FAGER, José	155
FERREIRA, Ana Gabriela Clipes	177
FLORES-VARGAS, Xochitl	269
FONSECA, Teresa	113
FONTANS, Exequiel	155
FRANCA, Aline	67
FREITAS, Juliana Lazzarotto	295

G

GIRÁLDEZ-REYES, Raudel	327, 349
GLÄNZEL, Wolfgang	229
GOMES, Henriette Ferreira	39
GONZALES-AGUILAR, Audilio	253
GONZÁLEZ-CABRERA, Odonel	339
GONZALEZ-TERUEL, Aurora	133
GONZÁLEZ-VALIENTE, Carlos Luis	305
GUIMARÃES, Maria Cristina Soares	9

H

HERNÁNDEZ-GONZÁLEZ, Vicenc	113
----------------------------	-----

J

JIMÉNEZ, Rodrigo Sánchez	19
JOVE-DELTELL, Carme	113

K

KUBOTA, Flávio Issao	167
----------------------	-----

L

LEGAZ-ARRESE, Alejandro	113
-------------------------	-----

M

MALDINI, Juan	155
MARCHIORI, Patricia Zeni	27
MARGAIX-FONTESTAD, Lourdes	133
MENA-CHALCO, Jesús Pascual	239
MENDOZA-RUANO, Juan-José	125
MIGUEL, Paulo Augusto Cauchick	167
MOYA-ANEGÓN, Félix de	281
MUGNAINI, Rogério	239
MUÑOZ-GARCÍA, Mitzi Lizeth	269

N

NASCIMENTO, Bruna Silva do	295
----------------------------	-----

P

PÉREZ-ANGÓN, Miguel Ángel	269
PÉREZ-ARREORTÚA, Noé	327
PÉREZ-GÁLVEZ, Grisel	339
PÉREZ-PINO, Nancy María	339
PETROCCELLI, Patricia	155
PICCO, Paola	155
PINTO, Adilson Luiz	253
PLASENCIA-SALGUEIRO, Armando	315

R

RABAÇO, Marcelo Henrique Leoni
REVERTER-MASÍA, Joaquín
RODRIGUES, Mara Eliane Fonseca
RODRÍGUEZ-FONT, Reinaldo Javier
RUIZ-PÉREZ, Taimi

S

SANTOS, Ana Paula Lima dos
SANTOS, Raquel do Rosário
SARTORI, Simone
SCALVENZI, Lisiane
SILVA, Cícera Henrique da
SILVA, Leyde Klebia Rodrigues da
SILVEIRA, Naira Christoletti
SIMON, Lucía

T

9 TSUNODA, Denise Fukumi

27

U

349 URDICIÁIN, Blanca Gil

19

V

91 VARGAS-QUESADA, Benjamín

281

39 VÍLCHEZ-ROMÁN, Carlos

143

77 VILLASEÑOR-GARCÍA, Elio Atenógenes

349

Y

9 YAMAKAWA, Eduardo Kazumi

167

Z

155 ZACCA-GONZÁLEZ, Grisel

281

Índice de assuntos

A			
Aceite vegetal	327	Desigualdad social	125
América Latina	305	Difusión de la innovación	133
Análise bibliométrica	77	Disciplinas informativas	305
Análise de conteúdo	51	Disseminação da informação	203
Análisis de contenido	327	Documento de patente	339
Análisis semiótico	269	E	
Análisis de redes sociales - América del Sur	253	<i>EndNote</i>	167
Avaliação da Pesquisa	239	Evaluación científica	281
Avaliação de desempenho	77	F	
Arquivo	51	Fontes de informação	177
Astronomia	269	Fontes de informação na Web	203
Atributos de la innovación	133	G	
B			
Bases de datos	315	Gerenciamento bibliográfico	167
Betim	51	Gestão ambiental	77
Bibliografias citadas	269	H	
Bibliometria	229, 253	<i>H-index</i>	143
Bibliometrics indicators	143, 229	Herramienta metodológica	125
Bibliometry	229	I	
Bibliotecas universitárias	39	Indicadores de patentes	339, 349
Biblioteconomia	91	Indicadores bibliométricos	113, 229, 281, 305
Boletines	113	Informação étnico-racial	203
Brasil	239	Information representation	27
Brecha digital	125	<i>Information sharing styles</i>	27
C		Interdisciplinaridade	91
Catalogação	67	Intertextualidad	269
Categorias de adoptantes	133	L	
Ciência da informação	91	Latindex	305
Cienciometria	281	Literatura indígena	67
Cientometria	253	Literatura periódica científica	295
Clasificación internacional de patentes	327, 349	M	
<i>Clustering</i>	19	<i>Marketing</i>	305
Colaboración científica	281	Mediação da informação	39
Combustión	327	<i>Mendeley</i>	167
Comércio eletrônico	191	México	269
Comparativo	167	Mídias sociais	9
Comportamiento informacional	133	Minería de textos	315
Comunicação científica	177	Modelo de espacio vectorial	19
Comunicación científica	155		
<i>Curriculum vitae</i>	155		
D			
Dengue	9		
Deporte	113		

Modelo generalizado de espacio vectorial	19	S	
Monitoramento de informação	9	Salud coletiva	281
Motores de búsqueda	315	<i>Scientific production</i>	143, 229
Movimento Negro da Paraíba	203	Scopus	113
N		Semântica latente	19
Navegação Web	191	Shale gas	349
O		Sistemas de organização do conhecimento	191
<i>OpenCourseWare</i>	133	Sistema Nacional de Investigadores	155
Organização do conhecimento	295	<i>Social bookmarking systems</i>	27
Organização da informação	191	<i>Social representation theory</i>	27
P		Sociedad de la Información	125
Patentes	327, 349	Sustentabilidade	77
Patrimônio imaterial	51	T	
Periódico científico	269	Taxonomia navegacional	191
Periódicos	177	Tendencias de investigación	305
Peruvian researchers	143	Tratamiento de agua	339
Pesquisa científica	295	Trayectorias tecnológicas	349
Produção científica	229, 239, 305	U	
Producción científica	113, 281	Uruguay	155
R		Uso da informação	203
Recuperación de información	315	V	
Redes sociais	9	Vigilância em saúde	9
Registro	51	Visibilidade	177
Registros bibliográficos	67	W	
Reinado de Nossa Senhora do Rosário	51	<i>Web of Science</i>	113
Referências Bibliográficas	167	<i>Web social</i>	39
Revistas científicas nacionais	239	Z	
		Zonas de Bradford	239
		Zotero	167

Agradecimentos

Acknowledgements

A revista *Transinformação* contou com a colaboração de especialistas *ad hoc* para a avaliação dos trabalhos a ela submetidos em 2014.

A

Adilson Luiz Pinto	UFSC
Adriana Bogliolo S. Duarte	UFMG
Adriane Maria Arantes de Carvalho	PUC/MG
Aldo de Albuquerque	IBICT
Amarildo Carnicel	PUC-Campinas
Ana Célia Rodrigues	UFF
Ana Isabel Bruzzi	USP
Ana Maria Almeida de Camargo	USP
Angélica conceição Dias Miranda	FURG
Armando Hirohumi Tanimoto	IFBA
Asa Fujino	USP

B

Brígida Maria Nogueira Cervantes	UEL
----------------------------------	-----

C

Camila Maciel Campolina Alves Mantovani	UFMG
Carlos Henrique Marcondes	UFF
Cladice Nobile Diniz	UNIRIO
Cibele Araújo Camargo Marques dos Santos	USP
Cibele Roberta Sugahara	PUC-Campinas

D

Daisy Pires Noronha	USP
Delsi Freis Davok	UDESC
Dulce Maria Baptista	UnB

E

Edna Lúcia da Silva	UFSC
Eduardo Amadeu Dutra Moresi	UCB
Elaine de Oliveira Lucas	UESC
Elisa Cristina Delfini Correa	UDESC
Ezequiel Theodoro da Silva	Unicamp

F

Fábio Botelho Josgrillberg	Universidade Metodista de São Paulo
Fábio Mascarenhas e Silva	UFPE
Fátima Maria Alencar Araripe	UFC
Fernanda Ribeiro Hierro	Universidade do Porto/Portugal
Fernando Cesar Lima Leite	UnB

Fernando Colen	UFMG
Frederico Vidigal	IBMEC/MG
Francisca Ferreira dos Santos	UFC

G

Giulia Crippa	USP
Guilherme Ataíde Dias	UFPB
Gustavo Henrique de Araújo Freire	UFPB

H

Helena de Fátima Nunes Silva	UFPR
Henriette Gomes	UFBA

I

Ida Regina Chitto Stumpf	UFRGS
Isa Maria Freire	UFPB
Ivette Kafure Münoz	UnB

J

Jacqueline Leta	UFRJ
João de Melo Maricato	UFG
Johanna Wilhelms Smith	USP
Jose Juan Peón Espantoso	UnB
José Maria Jardim	UFRJ
Juliano Serra Barreto	UnB
Jussara Borges de Lima	UFBA

L

Leandro Innocentini Lopes de Faria	UFSCar
Leilah Santiago Bufrem	UFPE
Lídia Alvarenga	UFMG
Lídia Silva de Freitas	UFF
Lígia Maria Arruda Café	UFSC
Lígia Maria Moreira Dumont	UFMG
Lucinéia Maria Bicalho	UFMG
Luis Fernando Sayão	CNEN

M

Marcelo Biassusi	EMATER/RS
Marcelo dos Santos	USP
Maria Celina Soares de Mello e Silva	MAST
Maria Cristiane Barbosa Galvão	USP
Maria das Graças Targino	UFPI
Maria de Fátima Tálamo	USP
Maria Inês Tomaél	UEL
Maria Luiza de Almeida Campos	UFF
Maria Nélide González de Gómez	IBICT
Maria Yêda Falcão Soares	UFBA
Marivalde Moacir Francelin	USP
Marta Lígia Pomim Valentim	Unesp
Miguel Angel Márdero Arellano	IBICT
Miriam Figueiredo Vieira da Cunha	UFSC
Monica Erichsen Nassif	UFMG
Murilo Bastos Cunha	UnB

N

Nádia Aurora Vanti Vitullo UFRN

O

Oswaldo Hajime Yamamoto UFRN

P

Patrícia Zeni Marchiori UFPR

Paulo de Martino Januzzi ENCE

R

Raimundo Nonato Macedo dos Santos UFPE

Raúl Quevedo-Blasco Universidade de Granada/Espanha

Regina Maria Marteleto UFRJ

Renato Kirchner PUC-Campinas

Renato Rocha Souza FGV

Ricardo Arcanjo de Lima Embrapa

Rogério Henrique de Araújo Júnior UnB

Rogério Mugnaini USP

Roniberto Morato do Amaral UFSCar

Rosa Inês de Novais Cordeiro UFF

Rosana Barros Boani Pauluci SCTIE

Rosângela Schwarz Rodrigues UFSC

Rose Marie Santini UFRJ

S

Sandra Miguel Universidade Nacional de la Plata/Argentina

Samile Andréa de Souza Vanz UFRGS

Silas Marques de Oliveira IASP

Silvana Drumond Monteiro UEL

Simone Cristina Dufloth FJP

Simone da Rocha Weitzel UNIRIO

Solange Maria Corder Unicamp

Sônia Elisa Caregnato UFRGS

Sônia Gondim UFBA

Sueli do Amaral UnB

Suzana Pinheiro Machado Mueller UnB

U

Úrsula Blattmann UFSC

V

Valéria Martins Valls FESP/SP

Vera Lúcia Dodebei UNIRIO

W

Waldomiro de Castro Santos USP

Wanda Aparecida Machado Hoffmann UFSCar

William Cruz UnB

Z

Zaira Regina Zafalon UFSCar

Instruções aos Autores

Transinformação é uma revista especializada, com periodicidade quadrimestral, aberta a contribuições da comunidade científica nacional e internacional, editada pela Faculdade de Biblioteconomia, Centro de Ciências Humanas e Sociais Aplicadas, Pontifícia Universidade Católica de Campinas. Fundada em 1989, é classificada na lista Qualis A1, publica artigos que contribuem para o estudo e o desenvolvimento científico da Ciência da Informação, da Biblioteconomia, da Arquivologia, da Museologia e de áreas afins.

Tipos de artigos aceitos

- *Original*: investigações baseadas em dados empíricos utilizando metodologia científica destinadas a divulgar resultados inéditos. Consistem de seções distintas: introdução, métodos, resultados, discussão e conclusão (limite máximo 5 mil palavras, preparados em espaço entrelinhas 1,5, com fonte *Arial* tamanho 11 e em folha formato A4).

- *Revisão (a convite)*: síntese crítica de tema de interesse da área, mediante análise e interpretação de bibliografia pertinente, de modo a conter uma análise crítica e comparativa dos trabalhos na área, que discuta os limites e alcances metodológicos, permitindo indicar perspectivas de continuidade de estudos naquela linha de pesquisa (limite máximo de 6 mil palavras, preparados em entrelinhas 1,5, com fonte *Arial* tamanho 11 e em folha formato A4).

- *Ensaio (a convite)*: reflexão sobre tema que origine futuras pesquisas (limite máximo de 5 mil palavras, preparados em entrelinhas 1,5, com fonte *Arial* tamanho 11 e em folha formato A4).

- *Comunicação*: relato de experiência ou de pesquisa acadêmica mediante o trabalho com um indivíduo, um grupo, uma comunidade ou organização, cuja divulgação possa contribuir para o melhor conhecimento dos recursos informacionais disponíveis e para a solução de problemas técnicos do setor (limite máximo de 5 mil palavras, preparados em entrelinhas 1,5, com fonte *Arial* tamanho 11 e em folha formato A4).

- *Tradução (reprodução) (a convite)*: artigos traduzidos, autorizados pelo detentor dos direitos de reprodução.

Envio de manuscritos

Todos os artigos devem ser submetidos de forma eletrônica pela página do Portal de Periódicos Científicos da PUC-Campinas <<http://www.puc-campinas.edu.br/periodicocientifico>>, cujos passos são os seguintes:

- Acessar o site <http://www.puc-campinas.edu.br/periodico_cientifico>.
- Escolher "Transinformação".
- Clicar em "Acessar a Revista".
- Já na página da Revista, entrar em "Acesso".
- Se for o primeiro acesso, preencher os dados pessoais no item "Cadastro". Se já estiver cadastrado, basta preencher *login* e senha.
- Para submeter trabalhos, siga as demais instruções do próprio sistema.

Manuscritos enviados por correio convencional, *fax*, *e-mail* ou qualquer outra forma de envio não serão apreciados pelos editores.

Processo de julgamento dos manuscritos

Todos os manuscritos inéditos só iniciarão o processo de tramitação se estiverem de acordo com as Instruções aos Autores. Caso contrário, **serão devolvidos para adequação às normas**, inclusão de carta ou de outros documentos eventualmente necessários.

Originais identificados com incorreções e/ou inadequações morfológicas ou sintáticas **serão devolvidos antes mesmo de serem submetidos à avaliação** quanto ao mérito do trabalho e à conveniência de sua publicação. Deve-se ainda evitar o uso da primeira pessoa "meu estudo..." ou da primeira pessoa do plural "percebemos...", pois em texto científico o discurso deve ser impessoal, sem juízo de valor e na terceira pessoa do singular.

Pré-análise: a avaliação é feita pelos Editores e Conselheiros com base na originalidade, pertinência, qualidade acadêmica e relevância do manuscrito para Ciência da Informação, Biblioteconomia, Arquivologia, Museologia e áreas afins. Os manuscritos podem ser rejeitados sem comentários detalhados após análise inicial, pelos Editores e Comissão Editorial da revista Transinformação, se os artigos forem considerados inadequados ou de prioridade científica insuficiente para publicação na Revista.

Aprovados nesta fase, os manuscritos serão encaminhados aos revisores *ad hoc* selecionados pelos editores. Cada manuscrito será enviado para dois revisores de reconhecida competência na temática abordada, podendo um deles ser escolhido a partir da indicação dos autores. Em caso de desacordo, o artigo será enviado para uma terceira avaliação.

O processo de avaliação por pares é o sistema *blind review*, procedimento sigiloso quanto à identidade tanto dos autores quanto dos revisores. Por isso os autores deverão empregar todos os meios possíveis para evitar a sua identificação.

Os pareceres dos revisores comportam três possibilidades: a) aprovação; b) recomendação de nova análise; c) recusa. Em quaisquer desses casos, o autor será comunicado.

Os pareceres são analisados pelos Editores que poderão aprovar ou não o manuscrito.

Manuscritos aceitos: manuscritos aceitos poderão retornar aos autores para ciência de eventuais alterações, no processo de editoração e normalização, de acordo com o estilo da Revista

Manuscritos recusados, mas com possibilidade de reformulação, poderão retornar como novo trabalho, iniciando outro processo de julgamento.

Conflito de interesse

No caso da identificação de conflito de interesse por parte dos revisores, o Comitê Editorial encaminhará o manuscrito a outro revisor *ad hoc*.

Forma e preparação de manuscritos

Procedimentos Editoriais

São aceitos trabalhos inéditos acompanhados de carta assinada por todos os autores, com identificação do tipo de artigo (Original / Revisão / Ensaio / Comunicação / Tradução), declaração de que o trabalho está sendo submetido apenas à revista *Transinformação* e cessão de direitos autorais.

A carta deve indicar o nome, endereço, números de telefone e e-mails dos autores e indicação do autor para o qual a correspondência deve ser enviada.

Os manuscritos submetidos, quando derivados de estudos que envolvem seres humanos, devem obrigatoriamente ter sido aprovados por **Comitê de Ética** em Pesquisa, conforme preconizam as diretrizes e normas da Resolução 196/96. Os autores deverão inserir a cópia digitalizada da declaração de aprovação do Comitê de Ética em pesquisa da Instituição.

Ao reproduzir no manuscrito material previamente publicado (incluindo textos, gráficos, tabelas, figuras ou quaisquer outros materiais), a **legislação cabível de direitos autorais** deverá ser respeitada e a fonte citada.

Opcionalmente, os autores podem indicar três possíveis revisores para o seu manuscrito.

Versão reformulada

O autor deverá enviar apenas a última versão reformulada do trabalho via site <<http://www.puc-campinas.edu.br/periodicocientifico>>. As modificações deverão ser sublinhadas, sendo anexada uma carta ao editor, reiterando o interesse em publicar nesta Revista e informando quais alterações foram processadas no manuscrito. Se houver discordância quanto às recomendações dos pareceristas, o autor deverá apresentar os argumentos que justificam sua posição. Caso o autor não encaminhe o manuscrito revisado e a carta-resposta no prazo estipulado, o processo editorial será encerrado, em qualquer etapa da submissão.

Provas: serão enviadas provas tipográficas aos autores para a correção de erros de impressão. As provas devem retornar ao Núcleo de Editoração na data estipulada. Outras mudanças no manuscrito não serão aceitas nesta fase.

Os manuscritos deverão apresentar

Página de título (única parte do texto com a identificação dos autores)

a) Título completo - deve ser conciso, evitando excesso de palavras, como "avaliação do..."; "considerações acerca de..."; "estudo exploratório...".

b) *Short title* com até quarenta caracteres (incluindo espaços), em português ou espanhol, inglês ou francês.

c) Nome de todos os autores por extenso, indicando a afiliação institucional de cada um. Será aceita uma única titulação e afiliação por autor. Os autores deverão, portanto, escolher, entre suas titulações e afiliações institucionais.

Observação: não havendo vínculo institucional, informar a atividade profissional, cidade e estado.

d) Todos os dados da titulação e da afiliação deverão ser apresentados por extenso, sem siglas.

e) Indicação dos endereços completos de todas as universidades às quais estão vinculados os autores.

f) Indicação de endereço para correspondência do autor responsável, quando mais de um, para a tramitação do original, incluindo fax, telefone e endereço eletrônico.

Outras observações

Resumo: todos os artigos submetidos em português, espanhol ou francês deverão ter resumo no idioma original e em inglês, com um mínimo de 150 palavras e máximo de 250 palavras.

Os artigos submetidos em inglês deverão vir acompanhados de resumo, título e palavras-chave em português e em inglês.

O resumo deve conter o objetivo do trabalho, os procedimentos metodológicos, os resultados, em determinados casos, e as conclusões. O texto não deve conter citações e siglas. Destacar no mínimo três e no máximo seis termos de indexação, utilizando os descritores em Ciência da Informação - Docutes - <http://eprints.rclis.org/bitstream/10760/5875/1/TESAURO_DOCUTES.pdf>.

Texto: com exceção dos manuscritos apresentados como Revisão, Ensaio e Comunicação, os trabalhos deverão seguir a estrutura formal para trabalhos científicos:

Introdução: deve conter revisão da literatura atualizada e pertinente ao tema, adequada à apresentação do problema, e que destaque sua relevância. Não deve ser extensa, a não ser em manuscritos submetidos como Artigo de Revisão.

Métodos: deve conter descrição clara e sucinta do método empregado incluindo as técnicas de pesquisa e os procedimentos adotados, e, se aplicável, o tratamento estatístico.

Em relação à análise estatística, os autores devem demonstrar que os procedimentos utilizados foram apropriados, contribuindo para a interpretação dos resultados obtidos.

Resultados: sempre que possível, devem ser apresentados em tabelas ou figuras, elaboradas de forma autoexplicativa, com análise estatística. Evitar repetir dados no texto.

Tabelas, quadros e figuras devem ter um título breve, ser limitados a cinco no conjunto, numerados com algarismos arábicos, consecutiva e independentemente, de acordo com a ordem de menção no texto. Enviar em folhas individuais e separadas, com indicação de sua localização no texto. É imprescindível a informação do local e ano do estudo. Os quadros e tabelas devem ter as bordas laterais abertas.

O autor se responsabiliza pela qualidade das figuras (desenhos, ilustrações, tabelas, quadros e gráficos), que deverão ser elaboradas em tamanhos de uma ou duas colunas (7 e 15cm, respectivamente); **não é permitido o formato paisagem**. Figuras digitalizadas deverão ter extensão jpeg e resolução mínima de 500 dpi.

Gráficos e desenhos deverão ser gerados em programas de desenho vetorial (*Microsoft Excel, CorelDraw X5, Adobe, Phoposhop CS6 SPSS v.10.0*), acompanhados de seus parâmetros quantitativos, em forma de tabela e com nome de todas as variáveis.

A publicação de imagens coloridas, após avaliação da viabilidade técnica de sua reprodução, será custeada pelo autor. Em caso de manifestação de interesse por parte do autor, a revista *Transinformação* providenciará um orçamento dos custos envolvidos, que poderão variar de acordo com o número de imagens, sua distribuição

em páginas diferentes e a publicação concomitante de material em cores por parte de outro autor.

Discussão: deve explorar, adequada e objetivamente, os resultados, discutidos à luz de outras observações já registradas na literatura.

Conclusão: apresentar as conclusões relevantes, considerando os objetivos do trabalho, e indicar formas de continuidade do estudo. **Não serão aceitas citações bibliográficas nesta seção.**

Siglas: deverão ser utilizadas de forma padronizada, restringindo-se apenas àquelas usadas convencionalmente ou sancionadas pelo uso, acompanhadas do significado, por extenso, quando da primeira citação no texto. Não devem ser usadas no título e no resumo.

Agradecimentos: podem ser registrados, em parágrafo não superior a três linhas, dirigidos a instituições ou indivíduos que prestaram efetiva colaboração para o trabalho.

Anexos: deverão ser incluídos apenas quando imprescindíveis à compreensão do texto. Caberá aos editores julgar a necessidade de sua publicação.

Citações no texto baseadas na NBR 10520/2002

Citações bibliográficas no texto: devem constar da lista de referências.

Não serão aceitas citações/referências de **monografias** de conclusão de curso de graduação e de **textos não publicados** (aulas, entre outros).

Se um trabalho não publicado, de autoria de um dos autores do manuscrito, for citado (ou seja, um artigo *in press*), será necessário incluir a carta de aceitação da revista que publicará o referido artigo.

Se dados não publicados, obtidos por outros pesquisadores, forem citados, será necessário incluir uma carta de autorização para uso dos mesmos.

Casos específicos

1) Citações literais de até três linhas: entre aspas, sem destaque em itálico e, em seguida, entre parênteses (Sobrenome do autor, data, página, sem espaço entre o ponto e o número). Ponto final depois dos parênteses.

2) Citações literais de mais de três linhas: em parágrafo destacado do texto, com 4cm de recuo à esquerda, em espaço simples, fonte menor que a utilizada no texto, sem aspas, sem itálico. Em seguida, entre parênteses: (Sobrenome do autor, data, página).

3) Vários autores citados em sequência: utilizar ordem cronológica de data de publicação dos documentos, separados por ponto e vírgula: (Crespo, 2005; Costa & Ramalho, 2008; Moresi *et al.*, 2010).

4) Textos com dois autores: Crippa e Bisoffi (2010) (no corpo do texto); (Crippa & Bisoffi, 2010) (dentro dos parênteses).

5) Textos com três ou mais autores: (Griselda *et al.*, 2009) (dentro dos parênteses) e Griselda *et al.* (2009) (fora dos parênteses).

6) Citações do mesmo autor publicados no mesmo ano: acrescenta-se letra minúscula após a data, sem espaçamento. Exemplo: (Morin, 2000a, 2000b).

Referências

As referências são baseadas na NBR-6023/2002. Recomenda-se limitar a 30 referências para artigos, exceto no caso de artigos de

revisão, que poderão apresentar em torno 50. Elas deverão ser ordenadas alfabeticamente pelo sobrenome do primeiro autor.

Casos específicos

1) Os títulos dos periódicos devem ser indicados por extenso.

2) Referências com autores e datas coincidentes usa-se o título do documento para ordenação e acrescenta-se letra minúscula após a data, sem espaçamento.

3) Referências com três ou mais autores, indica-se apenas o primeiro, acrescentando-se a expressão *et al.*

A exatidão e a adequação das referências a trabalhos que tenham sido consultados e mencionados no texto do artigo são de responsabilidade do autor.

Para outros exemplos recomendamos consultar as normas da ABNT-NBR-6023/2002.

Exemplos

Artigo com um autor

Oliveira, A. Direito à memória das comunidades tradicionais: organização de acervo nos terreiros de candomblé de Salvador, Bahia. *Ciência da Informação*, v.39, n.2, p.84-91, 2011.

Artigo com dois autores

Grippa, G.; Bisoffi, G.C. Memória e hipertexto: uma reflexão sobre o conhecimento relacional. *Transinformação*, v.22, n.3, p.233-246, 2009.

Artigo em suporte eletrônico

Moura, M.A. Informação e conhecimento em redes virtuais de cooperação científica: necessidades, ferramentas e usos. *DataGramaZero*, v.10, n.2, 2009. Disponível em: <<http://www.datagramazero.org.br>>. Acesso em: 16 maio 2009.

Livro

Grogan, D. *A prática do serviço de referência*. Brasília: Briquet de Lemos, 1995.

Livro em suporte eletrônico

Brasil. Ministério da Saúde. *Parto, aborto e puerpério: assistência humanizada à mulher*. Brasília: Ministério da Saúde, 2001. Disponível em: <<http://bibliotecadigital.puc-campinas.edu.br/services/e-books-MS/01-0420-M.pdf>>. Acesso em: 24 nov. 2010.

Capítulos de livros

Gasque, K.C.G.D. Teoria fundamentada: nova perspectiva à pesquisa exploratória. In: Mueller, S.P.M. (Org.). *Métodos para a pesquisa em ciência da informação*. Brasília: Thesaurus, 2007. p.107-142.

Capítulo de livro em suporte eletrônico

Fujita, M.S.L. O contexto da indexação para a catalogação de livros: uma introdução. In: Fujita, M.S.L. (Org.). *A indexação de livros: a percepção de catalogadores e usuários de bibliotecas universitárias*. São Paulo: Unesp, 2009. p.11-17. Disponível em: <http://www.esalq.usp.br/biblioteca/PDF/a_indexacao_de_livros_a_percepcao_de_catalogadores_e_usuarios_de_bibliotecas_universitarias.pdf>. Acesso em: 28 fev. 2013.

Dissertações e teses

Pereira, R. *Espaço Interativo (Ei!): o portal de relacionamento como suporte e estímulo à relação universidade-empresa*. 2009. Dissertação (Mestrado em Engenharia de Produção) - Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2009.

Trabalhos apresentados em congressos, seminários etc.

Amaral, M.S.; Pinho, J.A.G. Sociedade da informação e democracia: procurando a accountability em portais municipais da Bahia. In: Encontro da Associação Nacional de Pós-Graduação e Pesquisa em Administração, 32., 2008, Rio de Janeiro. *Anais...* Rio de Janeiro: EnANPAD, 2008. 1 CD-ROM.

Trabalhos apresentados em congressos, seminários etc. em formato eletrônico

Gauz, V.; Pinheiro, L.V.R. Fluxo da informação entre colecionadores, escribas e cientistas árabes na pré-institucionalização da ciência, séculos IV ao XV. In: Encontro Nacional de Pesquisa em Ciência da Informação, 11., 2010, Rio de Janeiro. *Anais eletrônicos...* Rio de Janeiro: Unirio, 2010. Disponível em: <<http://congresso.ibict.br/index.php/enancib/xianancib/paper/view/394/330>>. Acesso em: 20 abr. 2011.

Texto em formato eletrônico

Cieglinski A. *Bíblia é o livro mais lido e Monteiro Lobato o escritor mais admirado*. 2012. Disponível em: <http://www.ofaj.com.br/noticias_conteudo.php?cod=339>. Acesso em: 1 mar. 2013.

Lista de checagem

- Declaração de responsabilidade e transferência de direitos autorais assinada por cada autor.
- Verificar se estão completas as informações de legendas das figuras e tabelas.
- Preparar página de rosto com as informações solicitadas.
- Incluir o nome de agências financiadoras e o número do processo.
- Indicar se o artigo é baseado em tese/dissertação, colocando o título, o nome da instituição, o ano de defesa, em nota de rodapé.
- Categorizar os artigos em Original / Revisão / Ensaio / Comunicação / Tradução.
- Enviar a cópia do parecer do Comitê de Ética.
- Incluir título do manuscrito, em português, espanhol ou francês e em inglês.
- Incluir título abreviado (*short title*), com quarenta caracteres, para fins de legenda em todas as páginas.
- Verificar se as referências estão citadas no texto.
- Incluir permissão de editores para reprodução de figuras ou tabelas publicadas.

Documentos

Declaração de responsabilidade e transferência de direitos autorais

Cada autor deve ler e assinar os documentos (1) Declaração de Responsabilidade e (2) Transferência de Direitos Autorais, nos quais constarão:

- Título do manuscrito:

- Nome por extenso dos autores (na mesma ordem em que aparecem no manuscrito).

- Autor responsável pelas negociações:

1. Declaração de responsabilidade: todas as pessoas relacionadas como autoras devem assinar declarações de responsabilidade nos seguintes termos:

- "Certifico que participei da concepção do trabalho para tornar pública minha responsabilidade pelo seu conteúdo, que não omiti quaisquer ligações ou acordos de financiamento entre os autores e companhias que possam ter interesse na publicação deste artigo";

- "Certifico que o manuscrito é original e que o trabalho, em parte ou na íntegra, ou qualquer outro trabalho com conteúdo substancialmente similar, de minha autoria, não foi enviado a outra Revista e não o será, enquanto sua publicação estiver sendo considerada pela Transinformação, quer seja no formato impresso ou no eletrônico".

2. Transferência de Direitos Autorais: "Declaro que, em caso de aceitação do artigo, a revista Transinformação passa a ter os direitos autorais a ela referentes, que se tornarão propriedade exclusiva da Revista, vedada qualquer reprodução, total ou parcial, em qualquer outra parte, impressa ou eletrônica, sem que a prévia e necessária autorização seja solicitada e, se obtida, farei constar o competente agradecimento à Revista".

Assinatura do(s) autor(es)

Data ___ / ___ / ___

Justificativa do artigo

Destaco que a principal contribuição do estudo para a área em que se insere é a seguinte:

(Escreva um parágrafo justificando porque a revista deve publicar o seu artigo, destacando a sua relevância científica, a sua contribuição para as discussões, na área em que se insere o(s) ponto(s) que caracteriza(m) a sua originalidade e o conseqüente potencial de ser citado).

Dada a competência na área do estudo, indico o nome dos seguintes pesquisadores (três) que podem atuar como revisores do manuscrito. Declaro igualmente não haver qualquer conflito de interesses para esta indicação.

Revista Transinformação

Núcleo de Editoração SBI - *Campus I*

Rod. D. Pedro I, km 136 - Sala 8 - Prédio Antiga Reitoria - Pq. das Universidades - 13086-900 - Campinas - SP - Brasil

Fone/Fax: 55 (19) 3343-7401

E-mail: sbi.nucleoeditoracao@puc-campinas.edu.br

Home Page: <http://www.puc-campinas.edu.br/periodicocientifico/> / <http://www.scielo.br/tinf>

Guide for Authors

Transinformação is a specialized journal, issued three times a year, and open to contributions from the national and international scientific communities. It is edited by the School of Library Science, Center of Applied Human and Social Sciences, Pontifical Catholic University of *Campinas*. Founded in 1989, is classified in Qualis list A1, it publishes articles that contribute to the study and scientific development of Information Science, Library Science, Archival Science, Museology and related areas.

Types of articles accepted for publication

- *Original*: investigations based on empirical data using the scientific method that disclose original results. They are divided into distinct sections: introduction, methods, results, discussion and conclusion (number of words should not exceed 5 thousand with a line spacing of 1.5, Arial font size 11 and A4 paper size).

- *Review (only by invitation)*: critical synthesis of a theme of interest to the area based on the analysis and interpretation of the relevant literature containing a critical and comparative analysis of the works in the area, discussing the methodological limitations and scope, and indicating perspectives of more studies in the research line (number of words should not exceed 6 thousand with a line spacing of 1.5, Arial font size 11 and A4 paper size).

- *Essay (only by invitation)*: reflection about the theme that encourages new studies (should not exceed 5 thousand words with a line spacing of 1.5, Arial font size 11 and A4 paper size).

- *Communication*: report of an experiment or academic study of an individual, group, community or organization whose disclosure may contribute to better knowledge of the information available and to the solution of technical problems in the sector (should not exceed 5 thousand words with a line spacing of 1.5, Arial font size 11 and A4 paper size).

- *Translation (reproduction) (only by invitation)*: translated articles, authorized by the owner of the copyrights.

Manuscript submission

All articles must be submitted in electronic format at the Portal of Scientific Journals of *PUC-Campinas* at <<http://www.puc-campinas.edu.br/periodicocientifico>>, through the following steps:

- Access the site <<http://www.puc-campinas.edu.br/periodicocientifico>>
- Choose "*Transinformação*".
- Click on "Access the Journal".
- Once on the Journal's page, click on "Access."
- If this is your first visit, fill out your personal data in the item "Form". If you have already joined, just provide your login and password.
- To submit works, follow the instructions provided by the system.

Manuscripts sent by regular mail, fax, e-mail or other means will not be assessed by the editors.

Manuscript Review Process

All original manuscripts will only begin the review process if they conform to the Guide for Authors. Otherwise, they **will be returned to the authors for amendment** and inclusion of letter or other eventually necessary documents.

Manuscripts with mistakes and/or morphological or syntactic errors will be returned before they are submitted to assessment of their merit and convenience of publication. The authors must also avoid using the first person of the singular, "my study...", or the first person of the plural "we noticed...", since scientific texts ask for an impersonal, non-judgmental discourse.

Pre-analysis: the Editors and Counselors will assess the manuscript based on its originality, pertinence, academic quality and relevance to Information Science, Library Science, Archival Science, Museology and related areas. Submitted manuscripts may be rejected without detailed comments after initial review by at Editors and Editorial Board if the manuscripts are considered inappropriate or of insufficient scientific priority for publication in the Journal.

If the manuscript is approved in this phase, it will be sent to *ad hoc* referees selected by the editors. Each manuscript will be sent to two referees of known competence on the theme. One of them may be indicated by the authors. If the two referees disagree, the manuscript will be sent to a third referee for assessment.

Manuscript assessment relies on the double-blind review process, that is, the authors and referees remain mutually anonymous. Therefore, the authors are asked to do everything possible to avoid their identification.

The referees may give one of three opinions: a) approval; b) recommend a new analysis; and c) refusal. The author will always be informed of the referees' opinions.

The opinions will be analyzed by the Editors who will then approve the manuscript or not.

Accepted manuscripts may return to the authors for approval of possible changes in the editing and formatting process according to the style of the Journal.

Refused manuscripts that can be reformulated may be submitted to a new review process.

Conflict of interest

If the referees report conflict of interest, the Editorial Committee will send the manuscript to another *ad hoc* referee.

Manuscript layout and preparation

Editorial Procedures

The Journal accepts original articles accompanied by a letter signed by all authors identifying the type of article (Original / Review / Essay / Communication / Translation). The letter should also state that the manuscript is only being submitted to the Journal *Transinformação* and transfer the copyrights to the journal.

The letter should contain the name, address, telephone numbers and e-mails of the authors and indicate the address for correspondence.

The manuscripts submitted when derived from studies involving humans, are required to have been approved by the Research Ethics committee, as advocated in the guidelines and rules of Resolution 196/96. The authors should insert the scanned copy of the declaration of approval of the Research Ethics Committee.

Applicable copyright laws should be observed when previously published materials (texts, graphs, tables, figures or any others materials) are reproduced in the manuscript, including citation of the original source.

The authors may, instead, indicate three referees to assess their manuscript.

Reformulated version

The author is requested to send only the last reformulated version of the manuscript to the site <<http://www.puc-campinas.edu.br/periodicocientifico>>. Underline the changes made to the manuscript and a letter should be attached reiterating the author's interest in publishing in this Journal, indicating what changes were made to the manuscript. If the referees disagree, the author is asked to offer arguments that justify their position. If the author does not send the revised manuscript and the letter within the deadline, the editorial process will be terminated, regardless of the phase it is in.

Proofs: proofs will be sent to the authors for the correction of printing mistakes. Proofs must be returned to the Editorial Center by the stipulated deadline. Other changes to the manuscript will not be accepted during this phase.

The manuscripts should contain

Title page (only part of the text identifying the authors)

a) Full title – should be concise avoiding excess words, such as "assessment of...", "considerations about...", "exploratory study..."

b) Short title with at most forty characters (including spaces) in Portuguese or Spanish, and English or French.

c) Full name of all authors indicating the institutional affiliation of each one. Only one title and affiliation will be accepted per author. The authors must, therefore, choose among their titles and institutional affiliations.

Observation: if there is no institutional bond, please inform professional activity, city and state.

d) All title and affiliation data should be written in full, no abbreviations are accepted.

e) Include full addresses of all universities the authors have ties with.

f) Include full address of the corresponding author when there is more than one for correspondence regarding the manuscript, including fax and telephone numbers and e-mail(s).

Another observation

Abstract: all articles submitted in Portuguese, Spanish or French should have an abstract in the original language and one in English. The abstracts should have at least 150 words and at most 250 words.

The articles submitted in English should contain an abstract, title and keywords in Portuguese and in English.

The abstract should contain the objective of the study, the methodological procedures, the results in certain cases and the conclusions. The text should not contain citations and acronyms. Provide from 3 to 6 keywords using Information Sciences descriptors - Docutes - <http://eprints.rclis.org/bitstream/10760/5875/1/TESAURO_DOCUTES.pdf>.

Text: except for Reviews, Essays and Communications, all manuscripts must follow the formal structure for scientific studies:

Introduction: should contain a review of the current literature pertinent to the theme and appropriate to the presentation of the problem, emphasizing its relevance. It should not be extensive except for Reviews.

Methods: should contain a clear and compact description of the method used, including the study techniques and procedures, and, if applicable, the statistical treatment.

Regarding the statistical analysis, the authors should demonstrate that the procedures used were appropriate, contributing to the interpretation of the results.

Results: whenever possible, they should be presented in tables or figures, be self-explanatory and contain statistical analyses. Please avoid repeating data in the text.

Tables, charts and figures should have a brief title, total no more than five, and be numbered with Arabic numbers consecutively and independently according to the order in which they were mentioned in the text. They should be submitted in separate and individual sheets and the manuscript should indicate their location in the text. The location and year of the study must be included in their titles. Charts and tables should have open sides.

The author is responsible for figure quality (drawings, illustrations, tables, charts and graphs). Their sizes should occupy one or two columns (7 or 15cm wide, respectively). **The landscape format is not accepted.** Digital figures should be in jpeg format and have a minimum resolution of 500 dpi.

Graphs and drawings should be created in vector design software (Microsoft Excel, CorelDraw X5, Adobe, Phoposhop CS5 SPSS v.10.0), followed by their quantitative parameters in tables with the name of all variables.

The publication of color images after assessment of technical viability will be paid by the author. If the author request, the journal *Transformação* will provide a quote which may vary according to the number of images, image location (different pages), and the concomitant publication of color images by other author(s).

Discussion: should discuss the results in an appropriate and objective manner and compare them with other literature data.

Conclusion: present the relevant conclusions considering the objectives of the work and make suggestions for further investigations. **Bibliographical citations will not be accepted in this section.**

Acronyms: should be standardized and restricted to those used conventionally or sanctioned by use, followed by the full meaning the first time they appear in the text. They should not be used in the title and abstract.

Acknowledgments: may be done in a paragraph no longer than three lines and be directed to institutions or individuals who effectively collaborated with the study.

Attachments: should be included only when critical for the understanding of the text. The editors will decide if they should be published.

Text citations based on NBR 10520/2002

Bibliographical citations in the text should be included in the list of references.

Citations/references to undergraduate monographs and **unpublished texts** (classes, among others) **will not be accepted.**

If the unpublished work of one of the authors of the manuscript is cited (that is, an in-press article), the author must include the letter from the journal accepting to publish the article.

If unpublished data obtained by other researchers are cited, please include a letter authorizing its use.

Specific cases

1) Literal citations of up to three lines: in quotes, not in italic, followed by (Name of author, date, page, no space between period and number) in parentheses. The period should be placed after the closing parenthesis.

2) Literal citations with more than three lines must be in a separate paragraph, with a left margin of 4cm, using single space between the lines, smaller font than that of the text, without quotes, without italic, ending on the right margin of the text, followed by, in parenthesis, the last name of the author, date and page.

3) Many authors cited in sequence: use the chronological order in which the documents were published separated by a semicolon: (Crespo, 2005 & Costa; Ramalho, 2008; Moresi *et al.*, 2010).

4) Texts with two authors: Crippa and Bisoffi (2010) (in the body of the text); (Crippa & Bisoffi, 2010) (within parentheses).

5) Texts with three or more authors: (Griselda *et al.*, 2009) (within parentheses) and Griselda *et al.* (2009) (outside parentheses).

6) Citations of the same author published in the same year: add a lower-case letter after the date without spacing. Example: (Morin, 2000a, 2000b).

References

References are based on NBR-6023/2002. The references should be limited to 30 for articles but Reviews may have as many as 50. They should be ordered alphabetically according to the last name of the first author.

Specific cases

1) Journal titles should be written in full.

2) References with coinciding authors and dates should be ordered according to the title of the work and a lower-case letter is added after the date, without spacing.

3) In references with three or more authors, only the first author should be indicated followed by the expression *et al.*

The author are responsible for the accuracy and appropriateness of the references to studies that have been consulted and mentioned in the body of the manuscript.

For other examples, please refer to the ABNT-NBR-6023/2002 norms.

Examples

Article with one author

Ellis, D. A behavioural approach to information retrieval system design. *Journal of Documentation*, v.43, n.3, p.171-212, 1989.

Article with two authors

Xie, B.; Bugg, J. M. Public library computer training for older adults to access high-quality internet health information. *Library and Information Science Research*, v.31, n.3, p.155-162, 2009.

Article in electronic media

Willis, J.J.; Matrofski, S.D.; Weisburd, D. Making sense of COMSPAT: a theory-based analysis of organizational change in three police departments. *Law & Society Review*, v.41, n.1, p.147-188, 2007. Available from: <<http://onlinelibrary.wiley.com/doi/10.1111/j.1540-5893.2007.0294.x/abstract>>. Cited: Feb. 28, 2013.

Book

Allemang, D.; Hendler, J. Semantic web for the working ontologist: effective modeling in RDFS and OWL. 2nd.ed. Amsterdam: Morgan Kaufmann, 2011.

Book in electronic media

Lee, R.; Kim, H.K. (Ed.). *Computer and information science*. Berlin: Springer Berlin Heidelberg, 2008. Available from: <<http://link.springer.com/book/10.1007/978-3-540-79187-4/page/1#>>. Cited: Mar. 1, 2013.

Book chapters

Miller, D.H. User perception and the online catalogue: public library OPAC users think aloud. In: McIlwaine, I.A.C. (Ed.). *Knowledge organization and the global information society - ISKO*. London: Ergon Verlag, 2004. p.275-280.

Book chapter in electronic media

Copeland, B.J. Computation. In: Floridi, L. *The blackwell guide to the philosophy of computing and information*. Malden: Blackwell, 2004. p.1-17. Available from: <<http://onlinelibrary.wiley.com/book/10.1002/9780470757017>>. Cited: Feb. 28, 2013.

Dissertations and theses

Thompson, H.A. The significance and use of historical method in library and information science dissertations, 1984-1999. 2000. Dissertation (Master of Library and Information Science) - Kent State University, Kent, 2000.

Works presented in congresses, seminars etc.

Wang, X.; Liu, J. Usability evaluation of B2C web site. In: International Conference on Wireless Communications, Networking and Mobile Computing, 3., 2007, Shanghai. *Proceedings...* Washington: IEEE, 2007. p.3837-3840.

Works presented in congresses, seminars etc. in electronic media

Uskiv, V.; Uskova, M. Applications of telecommunications in education: National Science Foundation projects on advanced technological and

online education in information engineering technology. In: International Conference on Telecommunications, 10., 2003, Papeete. *Electronic proceedings...* Washington: IEEE, 2003. p.1701-1708. Available from: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1191689>>. Cited: Mar. 1, 2013.

Text in electronic media

Yen, D.H. *The Johari Window: The process of human interaction*. 1999. Available from: <http://www.noogenesis.com/game_theory/johari/johari_window.html>. Cited: Mar. 1, 2013.

Checklist

- Declaration of responsibility and transfer of the copyrights signed by each author.
- Verify if the information in the legends of figures and tables is complete.
- Prepare a cover page with the requested information.
- Include the name of the sponsors and process number.
- Indicate if the article is based on a thesis/dissertation and include in the footnote the title, name of institution and year of defense.
- Authors should indicate the article's category: Original / Review / Essay / Communication / Translation.
- Copy of the Research Ethics Committee approval.
- Include the title of the manuscript in Portuguese, Spanish or French and in English.
- Include the short title with a maximum of forty characters which will be placed on the footnote of all pages.
- Verify if all references are cited in the text.
- Include permission of editors for the reproduction of figures and tables that have been published elsewhere.

Documents

Declaration of responsibility and transfer of the copyrights

Each author should read and sign the documents (1) Declaration of Responsibility and (2) Transfer of Copyrights, which should include:

- Title of the manuscript:

- Full name of all authors (in the same order as they appear in the manuscript).

- Author responsible for the negotiations:

1. Declaration of responsibility: all people listed as authors should sign declarations of responsibility in the following terms:

- "I certify that I participated in the conception of the work and make public my responsibility for its content, and that I did not omit any financial relationships or agreements among the authors and companies that may benefit from the publication of this article".

- "I certify that the manuscript is original and the work, in part or in full, or any other work with substantially similar content of my authorship was not sent to another journal and will not be sent to another journal while its publication is being considered by *Transinformação*, whether in print or electronic format".

2. Transfer of the copyrights: "I declare that if the article is accepted for publication by the journal *Transinformação*, that the journal *Transinformação* will be the exclusive owner of its copyrights and any partial or full reproduction of the article anywhere else, in print or in electronic format, is forbidden unless previously authorized in writing by the abovementioned journal; if the authorization is granted, a statement will be added to the new article thanking the abovementioned Journal".

Signature of the author(s)

Date ____ / ____ / ____

Justification of the article

I emphasize that the main contribution of the study to the relevant area is the following:

(Write a paragraph justifying why the journal should publish your article, emphasizing its scientific relevance, contribution to the discussions in the relevant area affected by the aspect(s) of its originality and its consequent potential of being cited).

Given their competence in the study area, I indicate the name of the following three researchers that may act as referees of the manuscript. I also declare that there is no conflict of interests for this indication.

Journal Transinformação

Núcleo de Editoração SBI - Campus I

Rod. D. Pedro I, km 136 - Sala 8 - Prédio Antiga Reitoria - Pq. das Universidades - 13086-900 - Campinas - SP - Brasil

Fone/Fax: 55(19) 3343-7401

E-mail: sbi.nucleoeditoracao@puc-campinas.edu.br

Home Page: <http://www.puc-campinas.edu.br/periodicocientifico> / <http://www.scielo.br/tinf>

Pontifícia Universidade Católica de Campinas

(Sociedade Campineira de Educação e Instrução)

Grão-Chanceler: Dom Airton José dos Santos

Reitora: Profa. Dra. Angela de Mendonça Engelbrecht

Vice-Reitor: Prof. Dr. Germano Rigacci Júnior

Pró-Reitor de Graduação: Prof. Dr. Orandi Mina Falsarella

Pró-Reitora de Pesquisa e Pós-Graduação: Profa. Dra. Sueli do Carmo Bettine

Pró-Reitora de Extensão e Assuntos Comunitários: Profa. Dra. Vera Engler Cury

Pró-Reitor de Administração: Prof. Dr. Ricardo Pannain

Diretor do Centro de Ciências Humanas e Sociais Aplicadas: Prof. Me. Luís Arlindo Feriani Filho

Diretor-Adjunto do Centro de Ciências Humanas e Sociais Aplicadas: Prof. Me. Paulo Moacir G. Pozzebon

Transinformação

Com capa impressa no papel supremo 250g/m²
e miolo no papel couchê fosco 90g/m²

Capa / Cover

Kátia Harumi Terasaka

Editoração eletrônica / DTP

MRB Editoração

Impressão / Printing

E-Color Ltda

Tiragem / Edition

800

Distribuição / Distribution

Sistema de Bibliotecas e Informação da
PUC-Campinas - Serviço de Publicação,
Divulgação e Intercâmbio

Apresentação | *Presentation*

- 225 **TransInformação edita número temático - *Translformação edits thematic issue***
Rogério Eduardo Rodrigues Bazi
Seminário Internacional de Bibliometria - *International Seminar on Bibliometrics*
Jane M. Russell; Maria Victoria Guzmán; Isidro Aguillo; Francisco Collazo Reyes; Rogério Mugnaini

Artigos Originais | *Original Articles*

- 229 **Analysis of co-uthorship patterns at the individual level**
Análise dos padrões de coautoria a nível individual
Wolfgang Glänzel
- 239 **Comunicação científica no Brasil [1998-2012]: indexação, crescimento, fluxo e dispersão**
Scientific communication in Brazil [1998-2012]: indexing, growth, flow and dispersion
Rogério Mugnaini; Luciano Antonio Digiampietri; Jesús Pascual Mena-Chalco
- 253 **Visibilidad de los estudios en análisis de redes sociales en América del Sur: su evolución y métricas de 1990-2013**
Visibility of studies in social network analysis in south America: Its evolution and metrics from 1990 to 2013
Adilson Luiz Pinto; Audilio Gonzales-Aguilar
- 269 **Las prácticas de citación como interpretantes semióticos de acreditación de saberes locales en astronomía: México 1952-1972**
Citation practices as semiotic interpretants of certification of local knowledge in astronomy: Mexico 1952 - 1972
Francisco Collazo-Reyes; Xochitl Flores-Vargas; Mitzi Lizeth Muñoz-García; Miguel Ángel Pérez-Angón
- 281 **Producción científica cubana en Medicina y Salud Pública: Scopus 2003-2011**
Cuban scientific production in Medicine and Public Health: Scopus 2003 - 2011
Grisel Zacca-González; Benjamín Vargas-Quesada; Zaida Chinchilla-Rodríguez; Félix de Moya-Anegón
- 295 **A organização do conhecimento na dinâmica da pesquisa em artigos da literatura científica da Brapci**
Knowledge organization in the dynamics of research in the scientific literature of articles at Brapci
Juliana Lazzarotto Freitas; Bruna Silva do Nascimento; Leilah Santiago Bufrem
- 305 **Marketing en el campo de las disciplinas informativas: tendencias de investigación en América Latina [1985-2012]**
Marketing in the field of information disciplines: Research trends in Latin America [1985-2012]
Carlos Luis González-Valiente
- 315 **Análisis comparativo de herramientas de recuperación y análisis de información de acceso libre desde una concepción docente**
Comparative analysis of information retrieval and analysis of open acces tools from an education concept
Armando Plasencia-Salgueiro; Bárbara de los Milagros Ballagas-Flores
- 327 **Análisis de contenido del dominio tecnológico vegetable oil combustion**
Content analysis of the technological domain vegetable oil combustion
Noé Pérez-Arreortúa; Maidelyn Díaz-Pérez; Raudel Giráldez-Reyes; Humberto Andrés Carrillo-Calvet
- 339 **Estudio de patentes sobre tecnologías para tratamiento de agua y el agua residual**
A study of patenting on water and wastewater treatment technologies
Odonel González-Cabrera; Taimi Ruiz-Pérez; Mirelys Claro-Pérez; Nancy María Pérez-Pino; Grisel Pérez-Gálvez; Lisbey Collazo-Alfonso
- 349 **Tecnologías constituidas, innovaciones en proceso y tecnologías introducidas en el mercado internacional: caso de estudio**
Consolidated technologies ongoing innovations and technologies introduced to the international market: A case study
Maidelyn Díaz-Pérez; Raudel Giráldez-Reyes; Dayron Armas-Peña; Reinaldo Javier Rodríguez-Font; Elio Atenógenes Villaseñor-García; Humberto Andrés Carrillo-Calvet