



Transinformação

ISSN: 0103-3786

ISSN: 2318-0889

Pontifícia Universidade Católica de Campinas

BRITO, Aline Grasielle Cardoso de; QUONIAM, Luc; MENA-CHALCO, Jesús Pascual
Exploração da Plataforma Lattes por assunto: proposta de metodologia
Transinformação, vol. 28, núm. 1, 2016, Janeiro-Abril, pp. 77-86
Pontifícia Universidade Católica de Campinas

DOI: 10.1590/2318-08892016002800006

Disponível em: <http://www.redalyc.org/articulo.oa?id=384354425007>

- Como citar este artigo
- Número completo
- Mais informações do artigo
- Site da revista em redalyc.org

UABM
redalyc.org

Sistema de Informação Científica Redalyc
Rede de Revistas Científicas da América Latina e do Caribe, Espanha e Portugal
Sem fins lucrativos acadêmica projeto, desenvolvido no âmbito da iniciativa
acesso aberto

Exploração da Plataforma Lattes por assunto: proposta de metodologia

Investigation of the Lattes Platform by subject: A methodology proposal

Aline Grasielle Cardoso de BRITO¹

Luc QUONIAM¹

Jesús Pascual MENA-CHALCO²

Resumo

A Plataforma Lattes, no Brasil, é um relevante sistema de informação curricular que permite o registro da produção científica e tecnológica de cada pesquisador, sendo utilizada principalmente para avaliações isoladas de grupos ou instituições. Isso dificulta, quando não inviabiliza, o mapeamento de lacunas ou polos de conhecimento em áreas diversas de pesquisa, como por exemplo, nas buscas por assunto. O principal objetivo foi apresentar uma expressão de busca e uma metodologia para identificação do corpo de conhecimento em uma área específica, a Nanotecnologia, um dos campos prioritários ao desenvolvimento nacional. De acordo com o objetivo, o artigo é considerado um estudo de caso exploratório com abordagem quantitativa. Os procedimentos podem ser replicados em diversas áreas do conhecimento e permitem o uso da Plataforma Lattes como fonte de informação temática para análise e compreensão das mesmas, como a Nanotecnologia.

Palavras-chave: Nanotecnologia. Plataforma Lattes. Sistema de informação.

Abstract

In Brazil, the Lattes Platform is a curricular information system that allows researchers to manage their scientific and technological production, but it has been primarily used for reviewing and researching specific groups or institutions. Thus, it becomes difficult, if not unfeasible, to properly map the gaps or areas of knowledge in different fields of research. The aim of this study was to present an advanced search expression and a methodology to identify the body of knowledge in a specific area, such as nanotechnology, which is currently one of the research priority areas for national development. According to the objectives of the study, this is a quantitative study and an exploratory case study. The procedures may be replicated in different fields of knowledge and enable the Lattes Platform to be used as a source of thematic information for analysis and understanding of these fields, such as nanotechnology.

Keywords: Nanotechnology. Lattes Platform. Information system.

Introdução

A ciência acadêmica, considerada uma instituição, embora não tenha uma constituição escrita, se comporta

em essência como uma organização social, formada por diversos grupos de pessoas que dependem de laços de confiança pessoal e institucional para alcançarem seus objetivos comuns (Ziman, 2000). Os investimentos em

¹ Universidade Federal de São Carlos, Centro de Educação e Ciências Humanas, Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade, Rod. Washington Luis, km 235, 13565-905, São Carlos, SP, Brasil. Correspondência para/Correspondence to: A.G.C. BRITO. E-mail: <aline@nit.ufscar.br>.

² Universidade Federal do ABC, Centro de Matemática, Computação e Cognição, Pós-Graduação em Ciência da Computação. Santo André, SP, Brasil.

Recebido em 24/10/2014, reapresentado em 13/3/2015 e aceito para publicação em 13/4/2015.

Ciência, Tecnologia e Inovação (CT&I) contribuem para o compartilhamento dos resultados de pesquisas entre os pesquisadores. Nessa perspectiva, a atividade científica se torna mais e mais sociabilizada uma vez que o cientista isolado dá lugar ao pesquisador inserido em grupos e, conseqüentemente, na comunidade científica, entrando em contato permanente com seus pares.

Dois fatores tornam-se essenciais para o atual desenvolvimento da ciência: as redes de pessoas, que se formam pelos grupos de pesquisadores; e o grande número de informações disponibilizadas pela rede mundial *www (Internet)*. “Embora a forma de organização social em redes tenha existido em outros tempos e espaços, o novo paradigma da tecnologia da informação fornece a base material para sua expansão penetrante em toda a estrutura social” (Castells, 2010, p.10).

Nesse sentido, o exponencial crescimento de fontes de informação na *Internet* cria a necessidade de ferramentas que facilitem as buscas e análises dos dados disponíveis. Como respostas ao acúmulo desse volume de informações, vários mecanismos de busca têm sido implementados nos últimos anos (Brin & Page, 1998; Boutet & Quoniam, 2012), cujos sistemas de informação, atualmente quase sempre formados por informações em ambiente *Web*, não são facilmente visualizados e compreendidos e ainda necessitam de análises e ferramentas específicas para tal. Essas ferramentas de análise e medição dos dados, denominadas bibliometria, cientometria e infometria, são temas típicos da Sociologia da Ciência e da Ciência da Informação, que estudam os dados científicos, quantificando e interpretando-os.

Nos últimos anos, um especial interesse quanto à extração de conhecimentos em base de dados tem sido revelado. Esse processo é conhecido como mineração de dados, por permitir a descoberta de novos conhecimentos que podem ser obtidos a partir do tratamento de conjuntos de dados disponíveis nos repositórios de produção científica, por exemplo, em bancos de dados de produções bibliográficas, orientações acadêmicas, projetos de pesquisa e diretórios de grupos de pesquisa. No Brasil, há um relevante sistema de informação que pode ser utilizado para mensurar a produção científica e tecnológica de um determinado grupo de pesquisadores, ou até mesmo uma área, a Plataforma Lattes, existente desde meados dos anos 1980.

A Plataforma Lattes vem sendo utilizada, principalmente, para avaliações isoladas de grupos ou instituições nos estudos bibliométricos já publicados (Digiampietri, 2012). Isso dificulta, quando não inviabiliza, o mapeamento de lacunas ou polos de conhecimento nas diversas áreas de pesquisa.

Este trabalho elegeu a Nanotecnologia como assunto da pesquisa por ser uma das áreas prioritárias para a ciência no Brasil. Outros campos elencados na agenda de pesquisa do Ministério da Ciência, Tecnologia e Inovação são: Aeroespacial, Agropecuária, Biotecnologia, Defesa, Energia, Fármacos e Saúde, Nuclear, Petróleo e Gás, Tecnologias da Informação e Comunicação (Brasil, 2014). A Nanotecnologia é prioritária e considerada uma área de ponta que pode propiciar o desenvolvimento de diversas indústrias, tornando-as mais competitivas. O governo brasileiro investiu, de 2004 a 2014, pouco mais de 440 milhões de reais em pesquisa e desenvolvimento nessa área (Instituto de Estudos Brasil Europa, 2013).

Hoje, no Brasil, os estudos sobre Nanotecnologia estão centrados em 26 redes temáticas e 16 Institutos Nacionais de Ciência e Tecnologia (INCT). Existem mais de 260 grupos de pesquisa atuando na área no país com aproximadamente 2 500 pesquisadores e mais de 3 000 estudantes (Brasil, 2014).

De acordo com Prahalad e Hamel (1990) pessoas e organizações de destaque, em relação aos seus pares ou concorrentes, possuem um conjunto de habilidades essenciais chamadas de corpo ou núcleo do conhecimento. Além disso, representam o núcleo de um setor ou área do saber, favorecendo uma concentração e uma aproximação entre aqueles que as detêm. Neste estudo, o corpo do conhecimento será compreendido como o conjunto de pesquisadores que representam o núcleo de uma área da ciência. Os pesquisadores que compõem o núcleo são os mais citados, os que possuem mais relação entre os membros deste núcleo, os que mais publicam e, portanto, que obtêm maior prestígio.

Diante desse contexto, apresenta-se o principal questionamento deste estudo, isto é, verificar se é possível utilizar a Plataforma Lattes como fonte de informação temática para análise e compreensão de áreas específicas do conhecimento, como a Nanotecnologia.

O objetivo principal, portanto, é apresentar uma expressão de busca e uma metodologia para identificação do corpo de conhecimento nessa área específica.

A Plataforma Lattes e a Web 2.0

A Plataforma Lattes é uma iniciativa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) com o propósito de realizar a integração de bases de currículos acadêmicos de instituições públicas e privadas em uma única plataforma. Os chamados Currículos Lattes são atualmente considerados um padrão brasileiro de avaliação, representando um histórico das atividades científicas, acadêmicas e profissionais de pesquisadores cadastrados. Essa base é caracterizada pela livre inserção de dados.

Tim O'Reilly é o criador do termo e para ele a Web 2.0 não possui limites definidos, mas sim, um "núcleo gravitacional" por ser encarada como uma plataforma na qual os próprios usuários intervêm no desenvolvimento de um processo de organização dos dados. Não se depende, aqui, de *softwares offline* (editores de texto, imagem ou vídeos), mas de aplicativos executados *online*, via programa navegador da *Internet* (Primo, 2007).

A Plataforma Lattes é uma base de dados pensada dentro dos moldes da Web 2.0, a qual, dentre muitos conceitos importantes, não possui fronteiras rígidas, mas um centro gravitacional. A *Internet* passou a estimular a chamada "arquitetura de participação", algo que não fazia parte da primeira fase dos serviços *online*. Considerada como a segunda geração da *Internet*, a Web 2.0 aparece como potencializadora das formas de publicação, compartilhamento e organização das informações *online*. A Plataforma Lattes, de acordo com o CNPq, já ultrapassou 3 milhões de currículos, sendo 6,35% de doutores, 10,85% de mestres, 27,69% de graduados, 16,18% de especialistas, 35,54% de outros níveis e 3,39% de currículos que não possuem a informação (Estatística..., 2014).

A Lei de Acesso à Informação nº 12.527, de 18 de novembro de 2011, criou mecanismos para tornar efetivo o direito previsto na Constituição de que "órgãos e entidades públicas devem divulgar informações de interesse coletivo, sobretudo por meio da *Internet*, salvo aquelas

cujas confidencialidade esteja prevista no texto legal" (Brasil, 2011, *online*). Nesse sentido, todos esses milhões de currículos encontram-se disponíveis por meio da *Internet* e em formato HTML, constituindo uma vantajosa oportunidade de análise da CT&I do país.

Possibilidades de buscas e idiomas na Plataforma Lattes

A Plataforma Lattes está em português, mas oferece ao pesquisador a opção de divulgar seu currículo em inglês. Além disso, aceita qualquer idioma para o preenchimento dos campos, todavia, publicações em inglês são crescentes nas áreas de ciências duras e biológicas. Por conta disso, as palavras-chave para a expressão de busca criada foram utilizadas em inglês e português, no entanto, tal decisão dependerá sempre das características da área tratada.

A Plataforma Lattes oferece uma interface para busca e recuperação dos dados dos currículos, contudo os procedimentos para criação de expressão de busca não são apresentados de modo satisfatório. O sistema oferece uma ajuda (documentação) *online* que apresenta os operadores de relação e suas respectivas aplicações em uma expressão de busca, mas nem todos são reconhecidos pelo sistema. Nessa documentação (Conselho Nacional de Desenvolvimento Científico e Tecnológico, 2009a) são apresentados como válidos os operadores de relação as palavras "and", "or", "not" e "near", e os caracteres (), _ &, /, !, ~, *, ? e \$. No Quadro 1 é possível observar o resultado dos testes verificados com os operadores mencionados, notando-se que mais da metade deles não são reconhecidos pela Plataforma Lattes.

Para uma expressão de busca sofisticada é essencial que a base de dados aceite todos os operadores booleanos, principalmente os que permitam a truncagem das palavras - normalmente representados pelos caracteres \$, ? ou *. Mas nenhum deles é descrito como truncador válido na Plataforma Lattes.

Algumas das limitações referentes à Plataforma Lattes dizem respeito ao fato da mesma ser atualizada pelos próprios pesquisadores, não fornecendo padronização nesse quesito, ou seja, em um conjunto de currículos pode haver aqueles atualizados na semana passada e outros há alguns anos. Entretanto, pesquisa-

Quadro 1. Operadores reconhecidos pela Plataforma Lattes.

Operadores	Reconhecidos pela Plataforma
AND	Liga dois ou mais termos, limitando a busca.
OR	Liga dois termos e reúne todos os documentos que incluam pelo menos um deles.
NOT	Buscará registros que contêm o termo de pesquisa que o precede, mas não o termo que o sucede.
NEAR	Encontra documentos contendo ambas as palavras especificadas ou frases contendo até dez palavras entre elas.
()	Servem para elaborar pesquisas ainda mais complexas, definindo operações menores dentro da expressão inteira. A busca funciona nesse caso considerando os parênteses como se fossem termos isolados, e depois os combina.
Não reconhecidos pela Plataforma	
&	Liga dois ou mais termos, limitando a busca.
	Liga dois termos e reúne todos os documentos que incluam pelo menos um deles.
!	Buscará registros que contêm o termo de pesquisa que o precede, mas não o termo que o sucede.
~	Encontra documentos contendo ambas as palavras especificadas ou frases contendo até dez palavras entre elas.
Não descrito na documentação de busca da Plataforma	
_	Usado para unir bitermos.
*	Funcionaria como coringa com qualquer número de caracteres.
?	Funcionaria como coringa de qualquer caractere isolado.
\$	Funcionaria como coringa localizado em qualquer parte do texto.

Fonte: Conselho Nacional de Desenvolvimento Científico e Tecnológico (2014).

dores ligados a programas de Pós-Graduação tendem a manter seus currículos atualizados em função das avaliações da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) ou para pleitear financiamentos de pesquisas e/ou bolsas junto às agências de financiamento. Além disso, é questionável o grau de confiabilidade das informações cadastradas pelos pesquisadores em seus próprios currículos, pois, como não há supervisão direta, podem fornecer dados falsos a respeito de suas trajetórias.

Outro ponto importante que merece destaque é o fato da Plataforma Lattes servir como fonte de informação para órgãos que avaliam o Sistema Nacional de Pós-Graduação do Brasil, como a CAPES do Ministério da Educação (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, 2014) e, também, para as agências de fomento e financiamento de pesquisas e/ou oferta de bolsas. Contudo, embora haja inúmeras críticas quanto ao uso da Plataforma Lattes como fonte de informação em pesquisas, não se pode desconsiderar o seu ineditismo como uma base de dados que reúne e integra informações acadêmicas, profissionais, institucionais e bibliográficas de pouco mais de 3 milhões de pesquisadores.

Métodos

O artigo é um estudo de caso de natureza empírica, pois se trata de uma análise aprofundada de um ou mais objetos (casos), que permitem o seu amplo e detalhado entendimento (Gil, 1996). Seu objetivo é aprofundar o conhecimento acerca de um problema não definido suficientemente (Mattar, 1996), visando estimular a compreensão, sugerir hipóteses e questões ou desenvolver a teoria.

O estudo de caso pode ser classificado, por seu conteúdo e objetivo tal, como exploratório, explanatório ou descritivo. A principal tendência em todos os tipos é esclarecer o motivo pelo qual uma decisão ou um conjunto delas foram tomadas, como foram executadas e quais os resultados alcançados (Yin, 2001).

Ao iniciar o presente estudo, pesquisou-se, em diversas fontes de informação, as palavras que pudessem descrever a área de Nanotecnologia (Gogotsi, 2006; Porter *et al.*, 2008; Universidade Federal de São Carlos, 2009; Meyer *et al.*, 2010; Glossary..., 2011; Nanodictionary, 2011; Arora *et al.*, 2013; Wiktionary, 2013). Após levantamento inicial, a lista com 117 termos (em inglês e

português) foi submetida à análise de um especialista na área específica, que os aprovou, sugerindo a inserção de outros três.

A partir de cada termo pertencente à lista apresentada no Quadro 2, elaborou-se blocos de palavras da seguinte forma:

Bloco de palavras 1: (TERMO1 singular inglês OR TERMO1 plural inglês OR TERMO1 singular português OR TERMO1 plural português).

Foram identificados 120 blocos de palavras com 464 termos no total ao apresentá-los em inglês/português, singular/plural. Alguns deles foram desconsiderados e/ou agrupados, pois não havia equivalente em inglês/português ou por serem cognatos. Feito isso para todos eles, buscou-se na base de dados de Currículos Lattes o número de recuperação de cada bloco de palavras individualmente. Após esse levantamento, foi possível ordenar de forma decrescente os blocos que apresentaram os maiores resultados.

Quadro 2. Lista de descritores que definem a área de Nanotecnologia.

1. <i>bionanotechnology</i>	41. <i>nanobatteries</i>	81. <i>nanophotonics</i>
2. <i>nanocylinders</i>	42. <i>nanobeads</i>	82. <i>nanoplasma</i>
3. <i>nanodevice</i>	43. <i>nanobeam</i>	83. <i>nanopolição</i>
4. <i>nanodiamond</i>	44. <i>nanobiology</i>	84. <i>nanopolymers</i>
5. <i>nanodiffraction</i>	45. <i>nanobiotecnologia</i>	85. <i>nanopore</i>
6. <i>nanodispersions</i>	46. <i>nanobodies</i>	86. <i>nanoporous</i>
7. <i>nanodomain</i>	47. <i>nanocapsule</i>	87. <i>nanopowder</i>
8. <i>nanodot</i>	48. <i>nanocapturer</i>	88. <i>nanoprecipitation</i>
9. <i>nanodragster</i>	49. <i>nanocarro</i>	89. <i>nanoprobe</i>
10. <i>nanoelectromechanical</i>	50. <i>nanocatalysis</i>	90. <i>nanoprecipitation</i>
11. <i>nanoelectronic</i>	51. <i>nanoceramics</i>	91. <i>nanoreator</i>
12. <i>nanoeulsion</i>	52. <i>nanochip</i>	92. <i>nanoribbon</i>
13. <i>nanoespuma</i>	53. <i>nanociência</i>	93. <i>nanorings</i>
14. <i>nanoestruturado</i>	54. <i>nanoclay</i>	94. <i>nanorobótica</i>
15. <i>nanofabrication</i>	55. <i>nanocluster</i>	95. <i>nanorod</i>
16. <i>nanofiber</i>	56. <i>nanocomponents</i>	96. <i>nanoscratch</i>
17. <i>nanofilm</i>	57. <i>nanocomposites</i>	97. <i>nanosensor</i>
18. <i>nanofluid</i>	58. <i>nanocone</i>	98. <i>nanosheets</i>
19. <i>nanoformulations</i>	59. <i>nanocontainers</i>	99. <i>nanoshell</i>
20. <i>nanofotônica</i>	60. <i>nanocore</i>	100. <i>nanosources</i>
21. <i>nanogel</i>	61. <i>nanocrystal</i>	101. <i>nanospheres</i>
22. <i>nanograined</i>	62. <i>nanocrystalline</i>	102. <i>nanospring</i>
23. <i>nanohorn</i>	63. <i>nanomanufacturing</i>	103. <i>nanostuctured_lipid_carriers</i>
24. <i>nanohybrids</i>	64. <i>nanomáquina</i>	104. <i>nanosurgery</i>
25. <i>nanoinprint_lithography</i>	65. <i>nanomaterial</i>	105. <i>nanosuspension</i>
26. <i>nanoinprinting</i>	66. <i>nanomechanical</i>	106. <i>nanosystem</i>
27. <i>nanoindentation</i>	67. <i>nanomechanics</i>	107. <i>nanotechnologies</i>
28. <i>nanolitografia</i>	68. <i>nanomechatronics</i>	108. <i>nanotextured</i>
29. <i>nanomachine</i>	69. <i>nanomedicine</i>	109. <i>nanotige</i>
30. <i>nanomachining</i>	70. <i>nanometrology</i>	110. <i>nanotitania</i>
31. <i>nanomagnetism</i>	71. <i>nanomontador</i>	111. <i>nanotomography</i>
32. <i>nanomanipulation</i>	72. <i>nanomotor</i>	112. <i>nanotox</i>
33. <i>carbon_nanostructure</i>	73. <i>nano-object</i>	113. <i>nanotoxicologia</i>
34. <i>dendrimer</i>	74. <i>nanoparticle</i>	114. <i>nanotube</i>
35. <i>diamondoid</i>	75. <i>nanoparticle_tracking_analysis</i>	115. <i>quantum_dot</i>
36. <i>fullerene</i>	76. <i>nanoparticulate</i>	116. <i>quantum_dots</i>
37. <i>graphene</i>	77. <i>nanopatterning</i>	117. <i>quantum_well</i>
38. <i>molecular_machine</i>	78. <i>nanopeapods</i>	118. <i>quantum_wells</i>
39. <i>nanobalance</i>	79. <i>nanopharmaceuticals</i>	119. <i>sol_gel</i>
40. <i>nanobarras</i>	80. <i>nanophase</i>	120. <i>ultrananocrystalline_diamond</i>

Fonte: Elaborada pelos autores (2014).

Posteriormente, foi construída uma expressão com todos esses termos ligados pelo operador booleano OR, ou seja, individualmente o BLOCO 1 foi somado ao BLOCO 2 e assim sucessivamente. Tal expressão foi utilizada para recuperação dos currículos dos pesquisadores doutores brasileiros e estrangeiros. Ao todo, recuperou-se 11 295 currículos, mas, devido à limitação do tamanho da URL (7 386 caracteres) de resultados da recuperação, foi preciso readequar a expressão de busca gerada. Isso porque a mesma era composta de 24 692 caracteres e, desse modo, não possibilitava a visualização dos resultados.

Constatou-se que o bloco de palavras associado à nanopartícula (e suas variações em inglês e português, plural e singular) permitiu obter o maior número de currículos (ao todo 4 267). No entanto, quando os termos foram buscados de forma individual, obteve-se o resultado seguinte: *nanoparticle* (729), *nanoparticles* (3 130), *nanopartícula* (264) e *nanopartículas* (2 827).

Nesse sentido, para reduzir o tamanho da sequência de caracteres, optou-se por aplicar o seguinte filtro: excluir todos os blocos e termos de busca que retornaram um número 3% menor de currículos com relação ao bloco de maior resultado de recuperação. Como este último permitiu obter 4 267 currículos, o limiar considerado para exclusão dos blocos e termos foi de 128. No total havia 120 blocos de expressões de busca, porém foram descartados 86 (aproximadamente 71,9%) mantendo-se apenas 34, os quais conformaram a expressão final de busca. Essa limiarização sistemática permitiu identificar 11 125 Currículos Lattes.

Após a finalização de tais procedimentos a expressão pôde ser utilizada, uma vez que gerou uma URL válida em relação ao número aceitável de caracteres. Assim, obtiveram-se as seguintes possibilidades quando do tratamento dos dados:

1ª) 11 154 pesquisadores doutores em geral;

2ª) 8 250 pesquisadores doutores com presença no Diretório de Grupos de Pesquisa;

3ª) 2 256 pesquisadores doutores com presença no Diretório de Grupos de Pesquisa e com bolsa produtividade de pesquisa (1A, 1B, 1C, 1D E 2).

Para definir os participantes mais ativos da área de Nanotecnologia, consideraram-se apenas os pesqui-

sadores com doutorado, pois são aqueles com maior experiência. Além disso, o indivíduo precisava estar presente em algum grupo de pesquisa formalmente registrado no CNPq, pois estudos isolados possuem cada vez menos peso na comunidade científica. O terceiro filtro possível por meio da interface da Plataforma Lattes, foi o de detentores de Bolsa Produtividade. A mesma é “destinada aos pesquisadores que se destacam entre seus pares, valorizando sua produção científica segundo critérios normativos, estabelecidos pelo CNPq, e específicos, pelos Comitês de Assessoramento do CNPq” (Conselho Nacional de Desenvolvimento Científico e Tecnológico, 2009b, *online*).

A coleta das informações foi realizada mediante a interface de busca da Plataforma Lattes na *Web*, onde utilizou-se a expressão criada e o *software ScriptLattes* como ferramentas necessárias para acessar as informações do Sistema Gerenciador de Banco de Dados (SGBD) do CNPq. Na Plataforma Lattes não é permitido escolher os campos onde se deseja realizar as buscas e, desse modo, obteve-se a recuperação de currículos que tinham um dos termos da expressão de busca em qualquer campo do currículo. Esse fato resultou em uma alta revocação (relacionada à exaustividade), mas baixa precisão (relacionada à especificidade) (Lancaster, 2004). Por conta disso, houve a necessidade da utilização dos filtros. Após determinar os filtros permitidos pela interface da Plataforma Lattes, utilizou-se o *software ScriptLattes* (Mena-Chalco, 2009) para extração dos dados dos currículos.

Os relatórios gerados pelo *ScriptLattes* foram considerados um Sistema de Recuperação da Informação (SRI) por fornecerem:

Representação das informações contidas nos documentos, usualmente através dos processos de indexação e descrição dos documentos; Armazenamento e gestão física e/ou lógica desses documentos e de suas representações; Recuperação das informações representadas e dos próprios documentos armazenados, de forma a satisfazer as necessidades de informação dos usuários (Souza, 2006, 163).

Mediante os resultados da extração, realizou-se o último procedimento de filtragem do conjunto de pesquisadores, com o auxílio do *software Gephi*, utilizado para visualização e exploração de todos os tipos de redes e de técnicas de análise das mesmas (Bastian *et al.*, 2008).

Resultados e Discussão

Mediante os resultados iniciais foi possível observar o quanto o número de currículos aumenta na medida em que são acrescentados os blocos de palavras na expressão. Conforme apresentado na Figura 1(a), o número crescente de currículos recuperados tende a uma estagnação linear. O resultado dessa etapa é uma curva de saturação e pode ser visualizado nas Figuras 1 (a) e (b) as quais apresentam, respectivamente, o gráfico de currículos acrescentados e a curva de evolução da recuperação.

A sequência dos passos para criação da expressão de busca foi apresentada na seção de métodos. O principal resultado do presente estudo foi a criação da expressão de busca, apresentado no Quadro 3.

Após a definição dessa expressão, os identificadores (ID) de cada currículo Lattes foram recuperados por meio da utilização de um *script* específico para extração de dados programado em Python (Rossum, 2014). Esse *script* gera como resultado uma lista de todos os ID dos titulares dos currículos. Para este estudo optou-se pela 3ª possibilidade, isto é, com 2 256 pesquisadores, que resultou em uma URL com 4 300 caracteres. É importante ressaltar que através da PL não é possível baixar os dados dos currículos de modo automático. Por esse motivo, foi utilizado o *software* livre *ScriptLattes* (Mena-Chalco & Cesar-Júnior, 2009) para coleta dos dados dos currículos.

Toda a produção bibliográfica desses 2 256 pesquisadores, sem limitações temporais, compreendeu: 147 324 artigos completos publicados em periódicos;

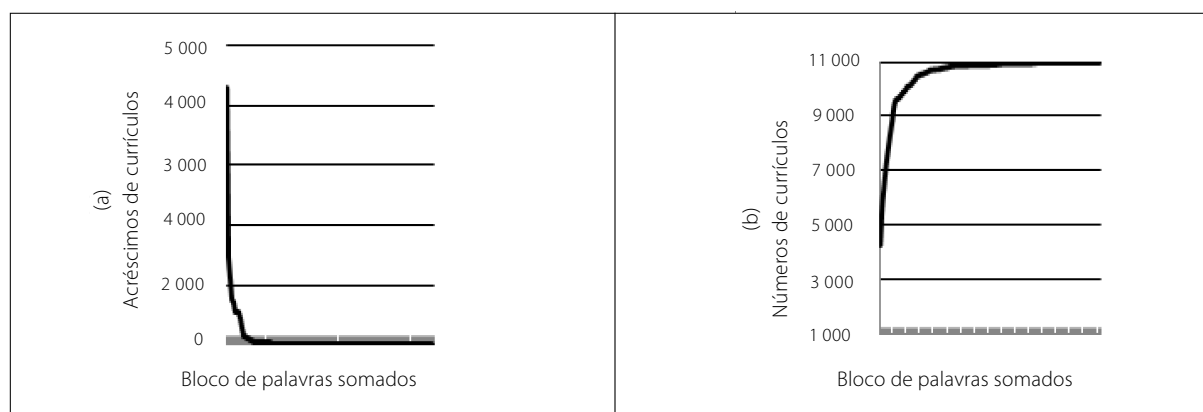


Figura 1. Representação das buscas.

Nota: (a) Curva de recuperação utilizando a interface da Plataforma Lattes. (b) Curva demonstrativa do número de currículos acrescentados.

Fonte: Elaborada pelos autores (2014).

Quadro 3. Expressão de busca bilíngue para Nanotecnologia na Plataforma Lattes.

(catalysts or fulereno or fulerenos or fullerene or fullerenes or grafeno or graphene or nanobiotecnologia or nanocapsulas or nanocapsules or nanociência or nanociências or nanocomposite or nanocomposites or nanocompósito or nanocompósitos or nanocristais or nanocristalino or nanocristalinos or nanocrystal or nanocrystalline or nanocrystals or nanoemulsão or nanoemulsions or nanoemulsões or nanoesferas or nanoestruturado or nanoestruturados or nanoestruturas_de_carbono or nanofiber or nanofibers or nanofibras or nanofios or nanoindentation or nanomateriais or nanomaterial or nanomaterials or nanoparticle or nanoparticles or nanopartícula or nanoparticulado or nanoparticulados or nanopartículas or nanoporous or nanopowders or nanoribbons or nanoscience or nanospheres or nanostructured or nanotechnology or Nanotecnologia or Nanotecnologias or nanotube or nanotubes or nanotubo or nanotubos or nanowire or nanowires or poços_quânticos or pontos_quânticos or quantum_dot or quantum_dots or quantum_well or quantum_wells or sol_gel).

Fonte: Elaborada pelos autores (2014).

2 728 livros publicados/organizados ou edições; 10 711 capítulos de livros publicados; 8 469 textos em jornais de notícias/revistas; 76 056 trabalhos completos publicados em anais de congressos; 18 559 resumos expandidos publicados em anais de congressos; 186 987 resumos publicados em anais de congressos; 2 561 artigos aceitos para publicação; 49 751 apresentações de trabalho e 4 823 demais tipos de produção bibliográfica, somando um total de 507 960 produções bibliográficas.

A partir da produção bibliográfica é possível compreender as atuações de coautoria ou colaboração, que ocorrem quando dois ou mais cientistas, trabalhando juntos em um projeto de pesquisa, compartilham recursos intelectuais, econômicos e/ou físicos. A coautoria é amplamente aceita no mundo científico e até estimulada por colocar diferentes autores, institutos e instituições em contato para enfrentar grandes problemas (Garcia *et al.*, 2010).

Com base nos conceitos de coautoria e auxílio das técnicas de análise de rede, utilizou-se mais um filtro, dessa vez por meio do *software* livre Gephi. A visualização das colaborações aponta os pesquisadores conectados à comunidade estudada. Foi utilizada a distribuição *Yifan Hu Proportional*, que possibilitou a visualização de uma rede simples apresentada na Figura 2 (a).

É possível observar um leve contorno distante do centro da rede. Esses nós periféricos são os pesquisadores desconectados do assunto abordado na expressão de busca, mas que foram recuperados pelos buscadores da Plataforma Lattes.

Em seguida, utilizou-se a estatística de “componente relacionado dirigido”, a qual detecta componentes forte e fracamente conectados, para eliminar os pesquisadores que não possuíam vínculos entre si, ou seja, com base no critério de coautoria, os excluídos por esse filtro são aqueles que não publicaram em parceria com os mais representativos da rede. Dessa forma, foram eliminados 67 pesquisadores, cuja rede (apresentada na Figura 2 (b)) apresentou 2 176 pesquisadores, Figura 2 (b). Ainda com esse mesmo objetivo, em um segundo momento, utilizou-se o filtro *K-Brace*, o que resultou na eliminação de mais 174 pesquisadores. A rede final é representada na Figura 2 (c) e possui 2 015 indivíduos.

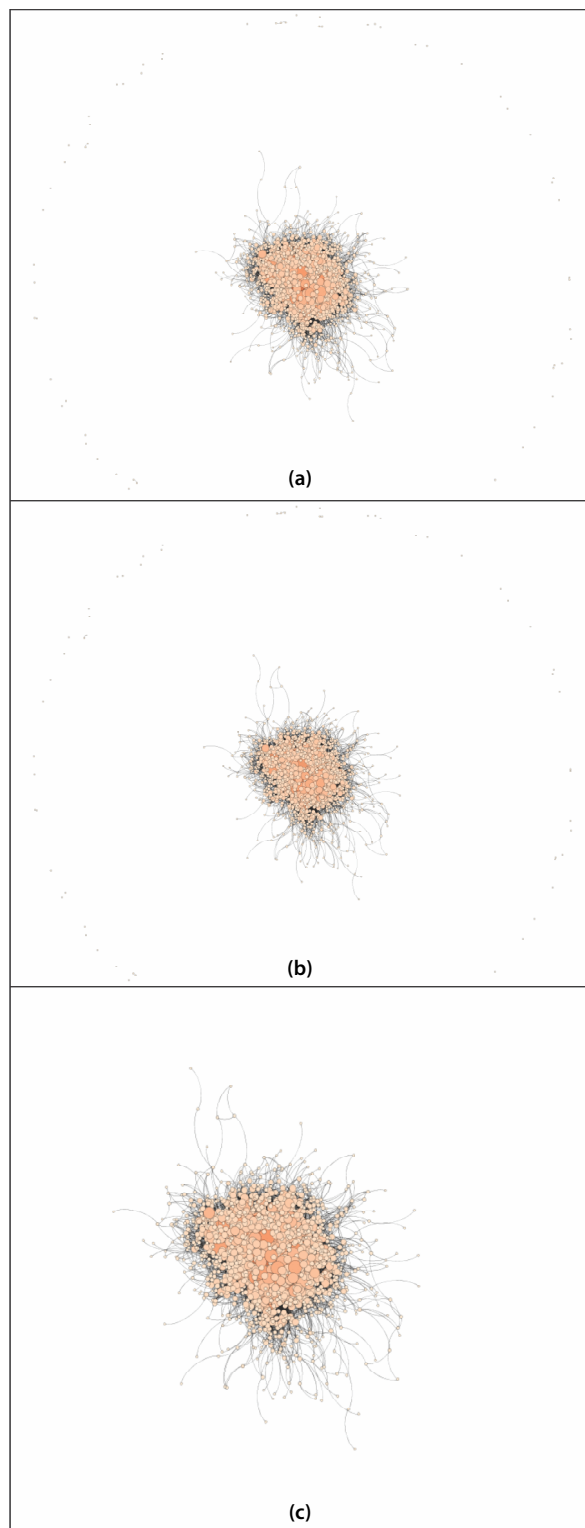


Figura 2. Visualização das redes em Nanotecnologia.

Fonte: Elaborada pelos autores (2014).

Nota: (a) Rede geral com 2 256 pesquisadores, (b) Rede com filtro Componente Relacionado e com 2 176 pesquisadores, (c) Rede com filtro *K-brace* e com 2 015 pesquisadores.

Conclusão

A identificação do corpo de conhecimento em áreas específicas, resultado deste trabalho, possibilita analisar diferentes aspectos da ciência brasileira, em uma perspectiva micro (pesquisador individual e grupos de pesquisadores) e macro (rede de colaboração ou representatividade da área). Tais análises são consideradas urgentes e poderão retratar o que é desenvolvido e publicado no Brasil em Ciência, Tecnologia e Inovação. Em contrapartida, nota-se forte concentração das análises bibliométricas clássicas, as quais se utilizam, convencionalmente, das bases de dados científicas internacionais. No entanto, justamente por serem internacionais, não representam a totalidade do que é produzido no Brasil, além de não contemplarem todas as áreas do conhecimento e nem possuírem o registro da trajetória dos pesquisadores. Esse é um positivo diferencial da Plataforma Lattes, pois possibilita a ampliação das fontes de informação, tais como: capítulos de livros, projetos em andamento, participação em bancas, orientações, prêmios recebidos.

A sistemática apresentada para criação de expressão de busca e curva de saturação na área de Nanotecnologia forneceu uma metodologia nova para identificação do corpo de conhecimento em uma área específica. O mesmo procedimento pode ser replicado em outros campos, respeitando suas especificidades em relação ao idioma e terminologia utilizada.

Diante dos resultados, constatou-se que a acessibilidade garantida pela Lei de Acesso à Informação não está plenamente abrangida no contexto da Plataforma Lattes, visto que o acesso aos dados dos currículos é permitido e disponibilizado, mas não admite uma manipulação efetiva e eficiente, exigindo um conjunto de técnicas específicas sobre recuperação e tratamento de dados. Isto é, requer conhecimentos que fogem ao domínio do usuário comum.

Colaboradores

Todos os autores contribuíram na concepção e desenho do estudo, análise de dados e redação final.

Referências

Arora, S.K. *et al.* Capturing new developments in an emerging technology: An updated search strategy for identifying nanotechnology research outputs. *Scientometrics*, v.95, n.1, p.351-370, 2013.

Bastian, M.; Heymann, S.; Jacomy, M. *Gephi, an open source graph visualization and manipulation software*. 2008. Available from: <<http://gephi.github.io/>>. Cited: Jun. 20, 2014.

Boutet, C.V.; Quoniam, L. Towards active SEO (Search Engine Optimization) 2.0. *Journal of Information System and Technology Management*, v.9, n.3, p.443-458, 2012.

Brasil. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. *Diário Oficial da União*, Brasília, DF, 18 set. 2011. Seção 1, (Edição extra). Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm>. Acesso em: 25 jun. 2014.

Brasil. Ministério da Ciência, Tecnologia e Inovação. *Programa nacional de nanotecnologia*. Brasília: MCT, 2014. Disponível em: <http://www.mct.gov.br/index.php/content/view/27137/Programa_Nacional_de_Nanotecnologia.html>. Acesso em: 11 set. 2014.

Brin, S.; Page, L. The anatomy of a large-scale hypertextual web search engine. Seventh International World Wide Web Conference. 7., Brisbane, 1998. *Proceedings electronics...* Brisbane: 1998. Available from: <<http://infolab.stanford.edu/pub/papers/google.pdf>>. Cited: 27 Sept. 2013

Castells, M. *A sociedade em rede*. São Paulo: Paz e Terra, 2010. v.1.

Conselho Nacional de Desenvolvimento Científico e Tecnológico. *Booleana: Wiki CNPq ajuda*. Brasília: CNPq, 2009a. Disponível em: <<http://ajuda.cnpq.br/index.php/Booleana>>. Acesso em: 21 set. 2014.

Conselho Nacional de Desenvolvimento Científico e Tecnológico. *Edital produtividade em pesquisa*. Brasília: CNPq, 2009b. Disponível em: <<http://www.cnpq.br/documents/10157/5f43cefd-7a9a-4030-945e-4a0fa10a169a>>. Acesso em: 12 set. 2014.

Conselho Nacional de Desenvolvimento Científico e Tecnológico. *Plataforma Lattes*. Brasília: CNPq, 2014. Disponível em: <<http://buscatextual.cnpq.br/buscatextual/busca.do?metodo=apresentar>>. Acesso em: 20 jan. 2014.

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. *Plataforma Sucupira*. Brasília: Capes, 2014. Disponível em: <<http://www.capes.gov.br/avaliacao/plataforma-sucupira>>. Acesso em: 21 set. 2014.

Digiampietri, L. A. *et al.* Minerando e caracterizando dados dos currículos Lattes. In: Brazilian Workshop on Social Network Analysis and Mining, 32., Curitiba, 2012. *Anais eletrônicos...* Curitiba: Sociedade Brasileira de Computação, 2012. Disponível em: <<http://professor.ufabc.edu.br/~jesus.mena/publications/pdf/digiampietri2012bdLattes.pdf>>. Acesso em: 25 maio 2014.

Estatísticas da base de currículos da Plataforma Lattes. *Painel Lattes*. 2014. Disponível em: <<http://estatico.cnpq.br/painelLattes/>>. Acesso em: 10 dez. 2014.

Garcia, C.C. *et al.* Authorship for scientific papers: The new challenges. *Revista Brasileira de Cirurgia Cardiovascular*, v.25, n.4, p.559-567, 2010.

Gil, A.C. *Como elaborar projetos de pesquisa*. São Paulo: Atlas, 1996.

Glossary of nanotechnology and related terms. *Rusnano*. 2011. Available from: <<http://eng.thesaurus.rusnano.com/wiki/letter.php?char=A&lang=E>>. Cited: 29 Jan. 2014.

Gogotsi, Y. (Ed.). *Nanomaterials handbook*. New York: Taylor & Francis, 2006.

Instituto de Estudos Brasil Europa. *Governo vai investir R\$440 milhões em Nanotecnologia*. Brasil: Ministério da Ciência, Tecnologia e Inovação, 2013. Disponível em: <<http://www.ibe.usp.br/index.php/pt/noticias/313-governo-vai-investir-r-440-milhoes-em-Nanotecnologia>>. Acesso em: 20 out. 2014.

Lancaster, F.W. *Indexação e resumos: teoria e prática*. 2.ed. Brasília: Briquet de Lemos, 2004.

Mattar, F.N. *Pesquisa de marketing: metodologia e planejamento*. São Paulo: Atlas, 1996.

Mena-Chalco J.P.; Cesar-Júnior, R.M. ScriptLattes: An open-source knowledge extraction system from the Lattes platform. *Journal of the Brazilian Computer Society*, v.15, n.4, p.31-39, 2009.

Mena-Chalco, J.P. *ScriptLattes software: uma ferramenta para extração e visualização de conhecimento a partir de currículos*

lattes. São Paulo, 2009. Disponível em: <<http://scriptLattes.sourceforge.net/>>. Acesso em: 20 abr. 2014.

Meyer, M.; Debackere, K.; Glanzel, W. Can applied science be *good science*? Exploring the relationship between patent citations and citation impact in nanoscience. *Scientometrics*, v.85, n.2, p.527-539, 2010.

Nanodictionary. *Nanodic.com: all about nano*. 2011. Available from: <<http://www.nanodic.com/>>. Cited: Mar. 20, 2014.

Porter, A. L. *et al.* Refining search terms for nanotechnology. *Journal Nanoparticle Research*, v.10, n.5, p.715-728, 2008.

Prahalad, C.K.; Hamel, G. The core competence of the corporation. *Harvard Business Review*, v.68, p.78-90, 1990. Available from: <http://www.enterrasolutions.com/media/docs/2013/08/Prahalad_Hamel_1990.pdf>. Cited: Jun. 10, 2014.

Primo, A. O aspecto relacional das interações na Web 2.0. *W-Compós*, v.9, n.1, p.1-21, 2007.

Rossum, G.V. *Python software foundation*. [s.l.]. 2014. Available from: <<http://wiki.python.org.br/>>. Cited: Apr. 10, 2014.

Souza, R.R. Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. *Perspectivas em Ciência da Informação*, v.11, n.2, p.161-173, 2006.

Universidade Federal de São Carlos. *Dicionário piloto de nanociência e Nanotecnologia*. São Carlos, 2009. Disponível em: <<http://www.etermos.cnptia.embrapa.br/produtos/produto97.html#letrad>>. Acesso em: 20 mar. 2014.

Yin, R. K. *Estudo de caso: planejamento e método*. 2.ed. São Paulo: Bookman, 2001.

Ziman, J.M. *Real science: What it is and what it means*. Cambridge: Cambridge University Press, 2000.

Wiktionary. *Category: English words prefixed with nano*. 2013. Available from: <http://en.wiktionary.org/wiki/Category:English_wor>. Cited: Oct. 14, 2013.

