

A utilização da ferramenta computacional Scriptlattes para avaliação das competências em pesquisa no Brasil

The use of Scriptlattes computational tool for brazilian research competences evaluation

Renato Ribeiro Nogueira Ferraz

Universidade Nove de Julho, São Paulo, Brasil
renatoferraz@uninove.br

Luc Marie Quoniam

Universidade Nove de Julho, São Paulo, Brasil
mail@quoniam.com

Resumo

A plataforma Lattes corresponde a um cadastro nacional dos currículos de pesquisadores brasileiros, sendo de preenchimento obrigatório para todos os indivíduos vinculados à instituições de ensino e pesquisa no Brasil. O Scriptlattes é uma ferramenta computacional capaz de baixar os currículos de pesquisadores cadastrados na plataforma Lattes, extraindo informações acadêmicas e profissionais, criando relatórios de produção, gráficos de colaboração e internacionalização da pesquisa, e mapas geográficos de investigação. Neste trabalho utilizou-se o Scriptlattes para demonstrar a eficácia desta ferramenta na avaliação das potencialidades em pesquisa em determinado tema, e também na organização da produção científica de um determinado departamento acadêmico. Metodologia: O Scriptlattes foi alimentado com a lista dos Professores de um Programa de *Stricto sensu* em Administração e com a lista de pesquisadores em nanotecnologia no território brasileiro. Resultados e

Abstract

*The curricular Lattes platform corresponds to a national register of Brazilian curricula researchers, being mandatory for all individuals linked to education and research institutions in Brazil. The Scriptlattes is a computational tool able to download the curricula of registered researchers in the Lattes Platform, extracting academic and professional information, creating production reports, charts collaboration and internationalization of research, and geographic maps of research. In this work we used the Scriptlattes to demonstrate the effectiveness of this tool in assessing the potential for research in a particular topics, and also in the organization of scientific production of na academic department. Methodology: The Scriptlattes was fed with a list of *Stricto Sensu* in management department teachers and also with the list of nanotechnology researchers in Brazil. Results and Conclusion: It was demonstrated the effectiveness of the tool in to the analysis and availability of the*

Conclusão: Demonstrou-se a eficácia da ferramenta no *information described above. The Scriptlattes proved to* tangente à análise e disponibilização das informações *be an effective way to obtain the lifting of meaningful* acima descritas. O Scriptlattes mostrou-se uma forma *performance indicators related to researchers and* eficaz para obtenção do levantamento de indicadores *research groups for a particular institutional or subject* significativos de desempenho relacionados aos *department, analyzing the progress of scientific* pesquisadores e aos grupos de pesquisa de um *production and contributing to know the research profile* determinado departamento institucional ou assunto, *results will be available in the coming years.* analisando o andamento das produções científicas e contribuindo para traçar o perfil das pesquisas cujos resultados estarão disponíveis nos próximos anos.

Palavras-chave: Scriptlattes. Stricto Sensu. **Keywords:** Scriptlattes. Sensu stricto. Lattes Platform. Plataforma Lattes. Nanotecnologia. *Nanotechnology.*

1. Introdução

O CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) desenvolveu uma plataforma curricular, a plataforma Lattes, com o intuito de unificar o cadastro nacional de pesquisadores no Brasil (Guedes, 2001). À partir de então o Brasil passou a contar com uma base nacional e única, contendo mais de 2 milhões de currículos (Alves, Yanasse, & Soma, 2012), capaz de qualificar os pesquisadores e permitir a viabilização do incentivo ao desenvolvimento científico e tecnológico nacional. Até o presente momento, não existe no mundo uma base curricular única semelhante à plataforma Lattes que, embora esteja hoje presente em diversos países da América Latina e também em Portugal, somente no Brasil encontra-se obrigatoriamente vinculada à concessão de fomento e à avaliação da qualidade do ensino superior no país.

Os currículos à disposição na base Lattes constituem, portanto, uma enorme fonte de informação não somente para análise individualizada, mas indubitavelmente, para avaliação de redes sociais estabelecidas entre os pesquisadores (Matheus, Parreiras, & Parreiras, 2006). No atual estágio da Ciência, ser um pesquisador isolado praticamente inviabiliza a execução de qualquer pesquisa, já que as associações entre pesquisadores e seus achados são essencialmente necessários para a produção do conhecimento científico (Balancieri, Bovo, Kern, Pacheco, & Barcia, 2005). Outra importante aplicação da plataforma refere-se à

possibilidade de acesso à informações que se encontram apenas nos currículos dos pesquisadores e não estão disponíveis nas bases referenciais, como por exemplo, os projetos submetidos, as orientações em andamento, dentre outras produções, permitindo que se possa traçar um panorama dos futuros resultados das pesquisas científicas.

O Scriptlattes, uma ferramenta de uso livre que trabalha em computadores pessoais que operem em ambiente Linux (já que o ambiente *Windows*[®] não possui todas as bibliotecas necessárias para sua operação), é capaz de baixar os currículos Lattes de um grupo de pesquisadores cadastrados no CNPq, extrair desses currículos as informações acadêmicas e profissionais que se julgarem necessárias, excluir produções redundantes e criar relatórios específicos que agrupem cada tipo de produção dos pesquisadores listados, além de gerar gráficos de colaboração, de internacionalização da pesquisa, e mapas geográficos de investigação relativos à lista de pesquisadores fornecida ao programa (Mena-Chalco, Junior, & Marcondes, 2009).

Com base na atual necessidade de se compilar de maneira organizada as produções acadêmicas e tecnológicas dos professores dos programas de *Stricto Sensu* no país com vistas à prestação de contas junto aos órgãos de avaliação da qualidade do ensino superior (Martins, Maccari, Storopoli, de Almeida, & Riccio, 2012), especialmente a CAPES, assim como aquelas relacionadas à avaliação das competências em pesquisa sobre um determinado assunto ou tema, e também relacionadas ao conhecimento das redes sociais estabelecidas entre os pesquisadores, o objetivo central deste trabalho foi demonstrar a utilidade da ferramenta computacional Scriptlattes na extração e disponibilização *on line* de toda a produção acadêmica gerada ou por uma lista de ID Lattes pré-determinada e criada manualmente (no caso da avaliação da produção acadêmica de um Programa de pós-graduação), ou por uma lista de ID Lattes gerada automaticamente pela plataforma Lattes em conjunto com o diretório de grupos de pesquisa do CNPQ (que neste caso referiu-se à lista de pesquisadores sobre o tema nanotecnologia no Brasil).

2. Metodologia

Faz-se agora uma breve descrição do mecanismo básico utilizado pela ferramenta Scriptlattes para a montagem das páginas no formato HTML. Informações mais detalhadas sobre o processo, assim como o passo-a-passo para a obtenção dos resultados aqui apresentados, poderão ser encontradas no trabalho original de Mena-Chalco e César Júnior (2009).

Embora não seja possível coletar dados diretamente da base Lattes, uma série de procedimentos computacionais permite que as informações dos currículos sejam recuperadas, com base em dados públicos disponibilizados pela própria plataforma. Cada pesquisador cadastrado na base Lattes possui um registro único constituído por um código de 16 algarismos, denominado ID Lattes, que dá acesso direto a uma página na *web* onde o currículo do referido pesquisador fica hospedado. Para a análise da produção de um departamento, ou mesmo para o levantamento das competências em pesquisa sobre determinado assunto, é necessário criar arquivos-texto que contenham as listas dos ID Lattes de todos os pesquisadores dos quais se queira extrair a produção. No caso da análise das produções científicas de determinado Programa ou departamento, tais listas são montadas manualmente, bastando para tal realizar uma procura pelos nomes dos pesquisadores na plataforma Lattes, anotando o número dos IDs. Quando se deseja utilizar a ferramenta para avaliar as competências por assunto, algumas etapas anteriores à preparação da lista de ID, listadas a seguir, devem ser observadas.

A primeira etapa de todo o processo, descrita recentemente (Quoniam, Jorge, André, & Jesús, 2013), gera um arquivo contendo uma lista dos ID Lattes de todos os pesquisadores que possuem uma palavra-chave qualquer em algum ponto de seus currículos Lattes. Especificamente no caso deste trabalho, foi gerada uma lista contendo os ID de todos os pesquisadores que possuíam a palavra-chave “nanotecnologia” tanto em seus currículos quanto registrada em algum Grupo de Pesquisa reconhecido pelo CNPQ dos quais os pesquisadores fossem integrantes. Vale lembrar que esse grupo restrito, denominado *core competence* em nanotecnologia, e composto por 1.915 pesquisadores na data de sua extração, originou-se de um tratamento anterior que utilizou como base uma lista de 12.237 currículos, proveniente de uma procura na plataforma Lattes utilizando-se como únicos

critérios o fato que os pesquisadores deveriam ser “doutores brasileiros e estrangeiros” e possuir a palavra-chave “nanotecnologia” registrada em qualquer ponto de seus Lattes.

Após a etapa anterior, a lista contendo os 1.915 ID Lattes do *core competence* em nanotecnologia alimentou o programa *Scriptlattes* versão 8.09-re08, a mais recente na data da extração. Finalizada a extração, obteve-se uma lista contendo 4.288 IDs de colaboradores considerados de nível 1 e denominados *core competence 1*. O critério utilizado para seleção e extração dos ID Lattes desses pesquisadores baseou-se na existência da palavra-chave “nanotecnologia” em seus currículos e também no fato de que tais pesquisadores deveriam ter alguma produção científica em colaboração com os 1.915 pesquisadores do *core competence*. Nesse caso, foram consideradas apenas as produções científicas de maior relevância acadêmica e pontuadas a seguir: “artigo completo publicado em periódico”, “artigo aceito para publicação” e “trabalho completo publicado em anais de eventos”. Dando continuidade ao trabalho, foi montada manualmente uma lista contendo os nomes e os IDs Lattes de todos os Professores do Departamento de Pós-graduação em Administração da Universidade Nove de Julho – UNINOVE – SP. Após a criação e alimentação do *Scriptlattes* tanto com as listas manuais para a avaliação da produção acadêmica do *Stricto Sensu* citado, quanto com as listas geradas automaticamente para avaliação das competências em pesquisa em nanotecnologia no país, o funcionamento do sistema dividiu-se em seis módulos sequenciais.

O primeiro módulo de funcionamento realiza o *download* dos currículos no formato HTML diretamente da plataforma Lattes, dedicando especial atenção à extração das informações relacionadas à publicação de artigos em periódicos. A segunda etapa do processamento busca na base Lattes as informações relacionadas ao endereço profissional, à lista de produções e à lista de supervisões em andamento e concluídas (limitadas ao intervalo de tempo fornecido), além da fotografia do pesquisador listado. Em seguida, ocorre a etapa de tratamento das redundâncias, onde o programa detecta e elimina as produções duplicadas após o pré-processamento dos currículos. Na quarta etapa a ferramenta *Scriptlattes* cria uma representação gráfica para demonstrar a rede de colaboração entre os membros listados de um grupo, baseando-se em suas produções científicas mais relevantes. Na penúltima fase do processamento é gerado um mapa de geolocalização mundial dos pesquisadores listados, baseado no endereço informado por eles à plataforma. Por fim, são

gerados todos os relatórios em formato HTML, relacionados às produções, bem como às orientações e supervisões em andamento e concluídas, separadas por tipo e por ano, e em ordem cronológica inversa (Mena-Chalco et al., 2009).

3. Resultados

A Figura 1A mostra a página em formato HTML gerada pela ferramenta Scriptlattes onde podem ser visualizados os *links* que dão acesso a diversas outras páginas, também no formato HTML, contendo as listas dos diferentes tipos de produção acadêmica sobre o tema nanotecnologia.

Figura 1A – Links de acesso às diferentes modalidades de produção científica em nanotecnologia no Brasil.

Análise da produção científica em nanotecnologia no Brasil	
[Membros Produção bibliográfica Produção técnica Produção artística Orientações Projetos Prêmios Eventos Mapa de geolocalização Internacionalização]	
Produção bibliográfica	
	<ul style="list-style-type: none">• Artigos completos publicados em periódicos (61286)• Livros publicados/organizados ou edições (974)• Capítulos de livros publicados (3444)• Textos em jornais de notícias/revistas (3622)• Trabalhos completos publicados em anais de congressos (42214)• Resumos expandidos publicados em anais de congressos (12920)• Resumos publicados em anais de congressos (88122)• Artigos aceitos para publicação (1499)• Apresentações de trabalho (22869)• Demais tipos de produção bibliográfica (2567)• Total de produção bibliográfica (239517)
Produção técnica	
	<ul style="list-style-type: none">• Produtos tecnológicos (2218)• Processos ou técnicas (2059)• Trabalhos técnicos (9466)• Demais tipos de produção técnica (7895)• Total de produção técnica (21638)

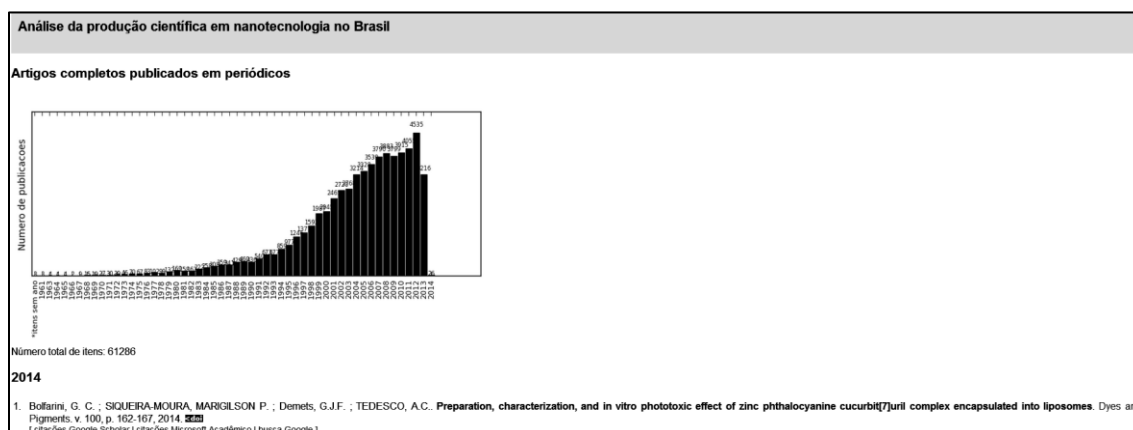
A Figura 1B mostra a página em formato HTML gerada pela ferramenta Scriptlattes onde podem ser visualizados os *links* que dão acesso às diversas listas de produção dos Professores do Programa de Pós-graduação avaliado neste trabalho, também no formato HTML.

Figura 1B - Links de acesso às diferentes modalidades de produção científica dos Professores do Programa de Stricto Sensu avaliado.

Todos os Professores Stricto Sensu em Administração - UNINOVE	
[Membros Produção bibliográfica Produção técnica Produção artística Orientações Projetos Prêmios Eventos Grafo de colaborações Mapa de geolocalização Internacionalização]	
Produção bibliográfica	
<ul style="list-style-type: none"> • Artigos completos publicados em periódicos (1275) • Livros publicados/organizados ou edições (188) • Capítulos de livros publicados (352) • Textos em jornais de notícias/revistas (579) • Trabalhos completos publicados em anais de congressos (1983) • Resumos expandidos publicados em anais de congressos (37) • Resumos publicados em anais de congressos (895) • Artigos aceitos para publicação (69) • Apresentações de trabalho (1074) • Demais tipos de produção bibliográfica (213) • Total de produção bibliográfica (6725) 	
Produção técnica	
<ul style="list-style-type: none"> • Produtos tecnológicos (19) • Processos ou técnicas (35) • Trabalhos técnicos (2766) • Demais tipos de produção técnica (809) • Total de produção técnica (3629) 	

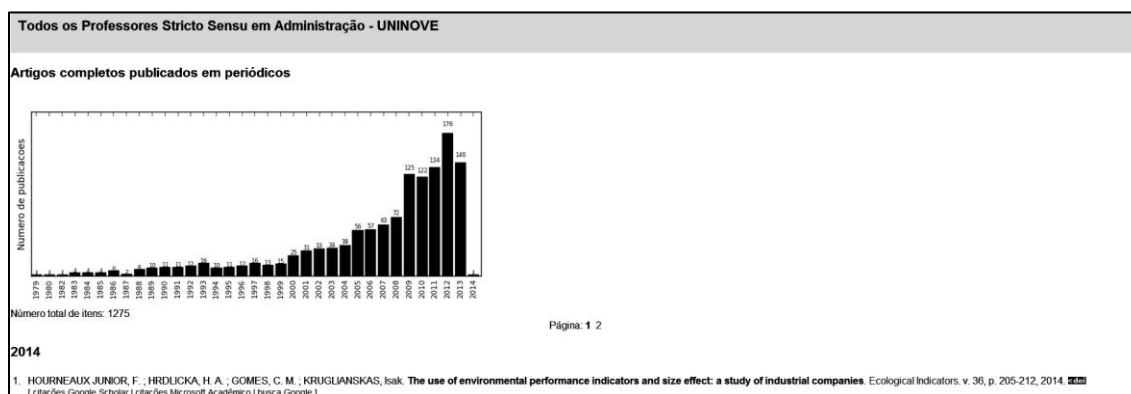
Ao acessar a primeira modalidade de produção bibliográfica disponível na Figura 1A (*Artigos completos publicados em periódicos*), obtém-se a lista completa destes artigos (total de 61.286 artigos sobre o tema nanotecnologia, publicados por diversos pesquisadores), organizados pelo ano de publicação. Na parte mais inferior também da Figura 2A pode ser observado o primeiro artigo publicado no ano de 2014 pelos pesquisadores do *core competence* em nanotecnologia. Para ter acesso direto ao *link* onde esses artigos estão hospedados, basta clicar na descrição do trabalho para que ocorra o direcionamento para a página de pesquisa do *Google Acadêmico* onde o artigo poderá ser acessado na íntegra ou parcialmente, dependendo da base de dados em que o mesmo estiver hospedado. Por fim, na parte superior da Figura 2A pode ser visualizado um gráfico de barras com a evolução numérica da produção científica sobre o tema nanotecnologia nas últimas cinco décadas.

Figura 2A - Gráfico com a crescente da produção científica sobre o tema nanotecnologia no Brasil nas últimas cinco décadas e link de acesso ao primeiro artigo completo publicado em periódico no ano de 2014.



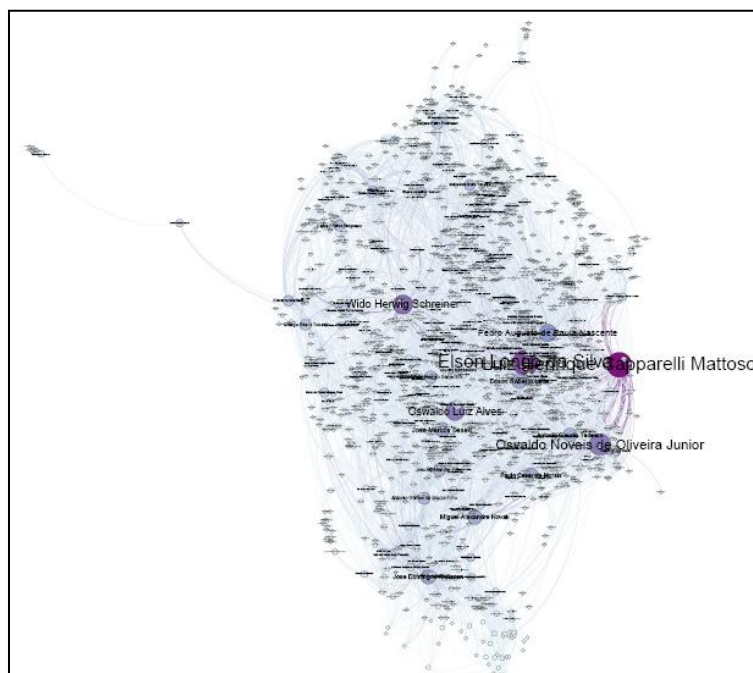
Da mesma forma, ao acessar a primeira modalidade de produção bibliográfica disponível na Figura 1B, é possível obter a lista completa dos artigos completos publicados em periódicos pelos diversos Professores do *Stricto Sensu* avaliado (total de 1.275 artigos no total) organizados pelo ano de publicação. Na parte mais inferior da Figura 2B pode ser observado o primeiro artigo publicado no ano de 2014 pelos Professores listados. O acesso à base de dados onde o artigo encontra-se hospedado dar-se-á ao clicar nas informações contidas na descrição do artigo. Por fim, na parte superior da Figura 2B também pode ser visualizado um gráfico de barras com a evolução numérica da produção científica nas últimas cinco décadas dos Pesquisadores associados ao *Stricto Sensu* avaliado.

Figura 2B – Gráfico com a crescente da produção científica nas últimas cinco décadas do Programa de Stricto Sensu avaliado e link de acesso ao primeiro artigo completo publicado em periódico no ano de 2014.



Como dito, a análise das relações estabelecidas pelos pesquisadores em um determinado assunto são essenciais para alavancar o conhecimento científico pertinente ao tema pesquisado (Freitas et al., 2008; Matheus et al., 2006). Atualmente, diversos *softwares*, boa parte deles de acesso livre e gratuito, encontram-se disponíveis para realização da análise das redes sociais de colaboração entre esses pesquisadores. Na Figura 3 observa-se uma rede tridimensional contendo os nomes dos pesquisadores pertencentes ao *core competence* em nanotecnologia no país, representados por esferas coloridas, rede essa preparada com a utilização do programa *Gephi* (Bastian, Heymann, & Jacomy, 2009). Na imagem podem ser vistas esferas de diâmetro maior, o que significa que o pesquisador por ela representado possui muitas colaborações com os outros pesquisadores do tema, colaborações essas indicadas pelas linhas contínuas que interligam as esferas. Quanto menor o diâmetro da esfera, menor é o número de colaborações efetuadas pelo pesquisador por ela representado.

Figura 3 – Análise da rede de colaborações entre os pesquisadores em nanotecnologia no Brasil com a utilização do software Gephi.



A análise da rede de colaborações entre os Professores do Programa avaliado feita também com a utilização do *software Gephi* (Figura 4) produziu um diferente modelo de rede tridimensional, também baseado na representação dos pesquisadores por esferas coloridas. Assim como no modelo anterior, o tamanho das esferas é diretamente proporcional ao número de relações que o pesquisador por ela representado possui com os demais pesquisadores do Programa de Pós-graduação avaliado, cujas esferas que os representam também estão unidas à esfera de maior volume do pesquisador citado através de linhas contínuas.

finalizadas, mas que também se tenha fácil acesso a informações como orientações e projetos em andamento conduzidos por cada pesquisador. A compilação e representação de forma transparente dessas informações com a utilização do Scriptlattes de permite que se possa traçar um panorama da futura produção acadêmica que contribuirá para o aumento do conhecimento científico ao longo dos anos, já que tais orientações e projetos de pesquisa em sua maioria resultarão em novas descobertas em curto e médio prazo, que darão origem à Dissertações, Teses, artigos e resumos em eventos e, especialmente, artigos completos publicados em periódicos.

5. Referências Bibliográficas

- Alves, A. D., Yanasse, H. H., & Soma, N. Y. (2012). LattesMiner: uma linguagem de domínio específico para extração automática de informações da Plataforma Lattes. Presented at the XII Workshop de Computação Aplicada, São José dos Campos. Retrieved from http://mtc-m18.sid.inpe.br/col/sid.inpe.br/mtc-m18/2013/01.15.16.10/doc/worcap2012_submission_61%20-%20Alexandre%20D.%20Alves.pdf
- Balancieri, R., Bovo, A. B., Kern, V. M., Pacheco, R. dos, & Barcia, R. M. (2005). A análise de redes de colaboração científica sob as novas tecnologias de informação e comunicação: um estudo na Plataforma Lattes. *Ciência Da Informação*, 34(1), 64–77.
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. In *ICWSM*. Retrieved from <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/viewPDFInterstitial/154Forum/1009>
- Freitas, C., Nedel, L. P., Galante, R., Lamb, L. C., Spritzer, A. S., Fujii, S., ... Moro, M. M. (2008). Extração de conhecimento e análise visual de redes sociais. *SEMISH-Seminário Integrado de Software E Hardware, Belém Do Pará, Brasil, SBC*, 106–120.
- Guedes, C. A. (2001). Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Currículo Lattes: Perguntas e Respostas. *Plataforma Lattes*. Retrieved November 6, 2013, from http://www.pucrs.campus2.br/manuais/dicas_lattes.pdf
- Martins, C. B., Maccari, E. A., Storopoli, J. E., de Almeida, M. I. R., & Riccio, E. L. (2012). A influência do sistema de avaliação nos programas de pós-graduação stricto sensu brasileiro. *Revista Gestão Universitária Na América Latina-GUAL*, 5(3), 155–178.

- Matheus, R. F., Parreiras, F. S., & Parreiras, T. A. S. (2006). Análise de redes sociais como metodologia de apoio para a discussão da interdisciplinaridade na ciência da informação. *Ci. Inf*, 35(1), 72–93.
- Mena-Chalco, J. P., Junior, C., & Marcondes, R. (2009). ScriptLattes: an open-source knowledge extraction system from the Lattes platform. *Journal of the Brazilian Computer Society*, 15(4), 31–39.
- Quoniam, L., Jorge, M., André, S., & Jesús, M.-C. (2013). Identificação de core competencies em dengue através da extração e tratamento de dados na plataforma brasileira Lattes (in press).