

# Um método orientado a propósito aplicado ao Currículo Lattes para fins de concessão de fomento a pesquisadores em grupos colaborativos

Kedma B. Duarte<sup>1</sup>, Rosina O. Weber<sup>2</sup> e Roberto C. S. Pacheco<sup>3</sup>

## Resumo

O desafio na concessão de fomento é determinar quais proponentes possuem maior chance de serem bem-sucedidos. Avaliações baseadas no mérito da proposta de pesquisa ou em bibliometria têm sido adotadas por gestores de ciência e tecnologia. No entanto, críticas a tais métodos questionam o fato de que os diferentes propósitos do fomento (pesquisa, formação e colaboração) não são considerados. Este trabalho apresenta um método orientado a específicos propósitos para mensurar a qualidade de pesquisadores. A abordagem é ilustrada com dados do Currículo Lattes, em um cenário de concessão de fomento a grupos colaborativos. Os resultados demonstram

## Abstract

*In research funding, the challenge is to determine which proposers are more likely to succeed. Assessments based on the research proposal merit, or bibliometrics, have been adopted by managers of science and technology. Criticisms to such methods question that the different purposes of assessments are not considered, for instance, the purposes of funding (research, scholarship and collaboration). Thus, this work presents a method to assess researcher quality oriented to specific purposes. The proposed approach is illustrated with data from the Curriculum Lattes in a scenario of research funding to research groups. The results demonstrate the value of the method, both in the understanding*

1 Doutora em Engenharia e Gestão do Conhecimento pela Universidade Federal de Santa Catarina (UFSC). Assessora técnico-científica da presidência da Fundação de Amparo à Pesquisa do Estado de Goiás (Fapeg).

2 Doutora em Engenharia de Produção pela UFSC. Professora associada na Drexel University, College of Computing & Informatics, Philadelphia, PA, USA.

3 Doutor em Engenharia de Produção pela UFSC. Professor adjunto no Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento na UFSC.

a usabilidade da solução proposta, tanto na compreensão do contexto dos grupos como na seleção de candidatos colaborativos com probabilidade de sucesso.

*of the context of the group, and in the selection of collaborative candidates likely to succeed.*

**Palavras-chave:** Mensuração da qualidade do pesquisador. Currículo Lattes. Fomento. Colaboração científica. Raciocínio Baseado em Casos. Aprendizagem de máquina

**Keywords:** *Researcher quality assessment. Curriculum Lattes. Funding. Research collaboration. Case-based reasoning. Machine learning*

## 1. Introdução

A concessão de fomento é uma tarefa que requer de gestores em ciência e tecnologia (C&T) a responsabilidade de avaliar a qualidade de pesquisadores. O desafio é determinar, nos diferentes propósitos de fomento (e.g., auxílio à pesquisa, formação ou colaboração), quais pesquisadores possuem maior chance de utilizá-lo de forma bem-sucedida.

O método de avaliação mais antigo e usado pelas agências de fomento, seja qual for o propósito, é a revisão por pares (SARTORI, 2011; MOLAS-GALLART, 2012), que consiste em solicitar a especialistas que julguem a significância da proposta de pesquisa (LAWANI, 1986). No entanto, devido a fatores como dependência e dificuldades em encontrar especialistas, a revisão por pares vem sendo substituída gradativamente por métricas baseadas na análise estatística de resultados científicos, isto é, na análise bibliométrica (LAWANI, 1986; HAUSTEIN; LARIVIÈRE, 2015; HICKS *et al.*, 2015). Clássicos exemplos são o índice de citações (GARFIELD, 1964), que mensura a qualidade de pesquisadores por meio de suas publicações (GARFIELD; MALIN, 1968); e a coautoria de artigos científicos (BEAVER, 1978), usado na mensuração da colaboração científica.

Porém, mesmo diante do avanço da bibliometria, o problema persiste e leva gestores de C&T a questionar a aplicação de métodos que não consideram a necessidade dos diferentes propósitos e contextos da avaliação (VAN NOORDEN, 2010; STI, 2016). Em resposta a tais criticismos, um grupo de pesquisadores propôs um conjunto de dez princípios sobre métricas de pesquisa denominado Manifesto de Leiden (HICKS *et al.*, 2015), que recomenda:

1. A avaliação quantitativa deve dar suporte à avaliação qualitativa especializada;
2. Medir o desempenho de acordo com a missão da instituição, do grupo ou do pesquisador;
3. Proteger a excelência da pesquisa localmente relevante;
4. Manter a coleta de dados e os processos analíticos abertos, transparentes e simples;
5. Permitir que os avaliados verifiquem os dados e as análises;
6. Considerar as diferenças entre áreas nas práticas de publicação e citação;
7. Basear a avaliação de pesquisadores individuais no juízo qualitativo da sua carreira;
8. Evitar solidez mal colocada e falsa precisão;
9. Reconhecer os efeitos sistêmicos da avaliação e dos indicadores;
10. Examinar e atualizar os indicadores regularmente (HICKS *et al.*, 2015, p.2-5).

No entanto, ainda não há na literatura uma metodologia que oriente a avaliação de pesquisadores de forma a seguir os princípios incluídos no Manifesto de Leiden (HICKS *et al.*, 2015).

O presente trabalho parte desse contexto, da busca por qualidade na concessão de fomento e do entendimento de que qualidade é *fitness-for-purpose* (JURAN; GODFREY, 1999), um conceito que a considera como dependente de perspectivas, necessidades e prioridades de usuários. Desta forma, a questão de pesquisa é: Como mensurar a qualidade de pesquisadores, de modo orientado ao propósito de concessão de fomento a grupos colaborativos?

Este trabalho apresenta, assim, um método orientado ao propósito da avaliação, para mensurar a qualidade de pesquisadores, que foi apresentado pelos autores deste trabalho em Duarte, Weber e Pacheco (2016) e detalhado em Duarte (2017). A abordagem segue os princípios do Manifesto de Leiden (HICKS *et al.*, 2015), sendo fundamentada em uma técnica de inteligência artificial chamada Raciocínio Baseado em Casos (RICHTER; WEBER, 2013) e faz uso de dados de currículos de pesquisadores.

O método proposto foi aplicado sobre dados do Currículo Lattes (*lattes.cnpq.br*), de forma a ilustrar sua usabilidade e simular um processo de concessão de fomento a grupos de pesquisadores. Os resultados demonstraram o potencial da solução para a compreensão do contexto de grupos colaborativos, assim como para a seleção de pesquisadores similares aos membros do grupo e, conseqüentemente, aptos a trabalharem em colaboração.

A solução contribui com gestores de Ciência, Tecnologia e Inovação (CT&I), entregando-lhes não somente um método, mas uma metodologia que permite, por meio de técnicas de inteligência artificial aplicadas em cada estágio, selecionar automaticamente pesquisadores aptos ao propósito da avaliação. Em particular, o estudo abre potenciais perspectivas sobre o uso do Currículo Lattes (*lattes.cnpq.br*) e do Diretório de Grupos de Pesquisa do Brasil (DGP) (*lattes.cnpq.br/web/dgp*) na compreensão, por exemplo, do contexto dos grupos colaborativos.

O presente estudo apresenta, na próxima seção, a fundamentação teórica do trabalho; e introduz, na seção 3, o método orientado a propósitos para mensurar a qualidade de pesquisadores (DUARTE, 2017). Na Seção 4, um estudo de caso ilustra a aplicação do método proposto, seguido, na Seção 5, das considerações finais e trabalhos futuros.

## 2. Referencial teórico

Nesta seção, são introduzidos conceitos e técnicas usadas no método proposto. Inicialmente é apresentado um breve relato sobre o *curriculum vitae* (CV), seu uso em estudos bibliométricos, citando como exemplo o Currículo Lattes (*lattes.cnpq.br*). Em seguida, são apresentadas técnicas baseadas em inteligência artificial adotadas na implementação do método proposto, como, por exemplo, a Aprendizagem de Máquina e Raciocínio Baseado em Casos (RBC).

### 2.1. *Curriculum vitae*

O *curriculum vitae* é um documento histórico no qual o próprio pesquisador registra sua trajetória de vida acadêmica, isto é, sua formação educacional, atuação profissional, produções científicas, colaborações e fomentos recebidos em projetos de pesquisa, dentre outros (DIETZ *et al.*, 2000).

A riqueza de conhecimento contida no currículo faz deste um atrativo e potencial fonte de conhecimento para estudos em bibliometria, principalmente sobre mobilidade de pesquisadores e colaboração internacional (GAUGHAN, 2009; SANDSTRÖM, 2009). Além disso, os currículos representam valiosas fontes de informação para indicadores de C&T em universidades, institutos de pesquisa e agências de fomento (CGEE, 2016; FURTADO *et al.*, 2015).

Uma das maiores referências de repositórios de dados de currículos é a base de dados do Currículo Lattes (*lattes.cnpq.br*). Criado em 1999 pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), o Currículo Lattes constitui o núcleo da Plataforma Lattes (PACHECO, 2003). A base Lattes foi reconhecida internacionalmente pela alta qualidade de seus dados (LANE, 2010) e conta, atualmente, com mais de cinco milhões<sup>4</sup> de currículos (CNPq, 2018).

---

4 Mais informações podem ser obtidas no sítio: <http://estatico.cnpq.br/painelLattes/mapa>

A plataforma contempla quatro unidades de informação: currículo; grupos de pesquisa; projetos de pesquisa; e instituições. O Currículo Lattes ([lattes.cnpq.br](http://lattes.cnpq.br)) e o Diretório de Grupos de Pesquisa do Brasil (DGP) ([lattes.cnpq.br/web/dgp](http://lattes.cnpq.br/web/dgp)) serviram como fonte de dados para este trabalho.

## 2.2. Aprendizagem de máquina

A aprendizagem de máquina (do inglês, *Machine Learning*) consiste, em linhas gerais, na criação de classificadores a partir da identificação de características (*i.e.*, atributos) comuns em um conjunto de exemplares de dados. Uma técnica popular que consiste em estudar a relativa relevância de cada atributo para fins de classificação é a aprendizagem de pesos (do inglês, *Feature Weighting*) (AHA, 1998; HALL, 1999). Essa técnica enfatiza a importância dos atributos pela alocação de pesos a cada um deles, produz a relevância relativa dos mesmos e, conseqüentemente, permite a sua ordenação pelos pesos atribuídos. Um exemplo desta técnica é o algoritmo *Relief*, desenvolvido por Kira e Rendell (1992), que identifica estatisticamente as características relevantes a um determinado problema. Outra solução baseada em aprendizagem de pesos é o algoritmo *Correlation Based Feature Selection* (CFS), proposto por Hall (1999), que avalia a relevância do atributo pela aplicação de medidas de correlação.

## 2.3. Raciocínio baseado em casos

Raciocínio baseado em casos (RBC), (do inglês, *Case-Based Reasoning*), é uma metodologia de inteligência artificial (do inglês, *Artificial Intelligence*) resultante de pesquisas em teorias da ciência cognitiva e ciência da computação (RICHTER; WEBER, 2013). O RBC combina aprendizagem de máquina e raciocínio, fundamentando-se no armazenamento de experiências passadas que possam ser adaptadas para ajudar a resolver problemas futuros (MÁNTARAS *et al.*, 2005).

Desta forma, para resolver um novo problema, uma solução é recuperada da memória, apropriadamente adaptada e aplicada ao novo problema, sendo as novas soluções armazenadas para uso futuro (CARBONELL; MICHALSKI; MITCHELL, 1983). Um ciclo de RBC proposto por Aamodt e Plaza (1994) consiste em quatro etapas básicas: recuperação (*retrieve*), reutilização (*reuse*), revisão (*revise*) e retenção (*retain*).

Este ciclo representa o núcleo dos métodos que adotam RBC como uma metodologia capaz de combinar, em cada etapa, diferentes técnicas e abordagens de aprendizado de máquina (AHMED; BEGUM; FUNK, 2012; WATSON, 1999); e ocorre da seguinte forma:

- Recuperação: Dado um problema (novo caso), a missão é recuperar um caso candidato da base de casos, de forma que este seja tão similar ao novo caso, que suas soluções sejam intercambiáveis. Para este objetivo, métricas são aplicadas para calcular o grau de similaridade entre os dois casos.
- Reutilização: a solução recuperada é reutilizada e adaptada ao novo problema.
- Revisão: fontes externas ou usuários verificam se a solução proposta realmente resolve o problema e permitem, assim, que os ajustes necessários sejam realizados.
- Retenção: etapa final, na qual a solução é retida na base de conhecimento para futura reutilização.

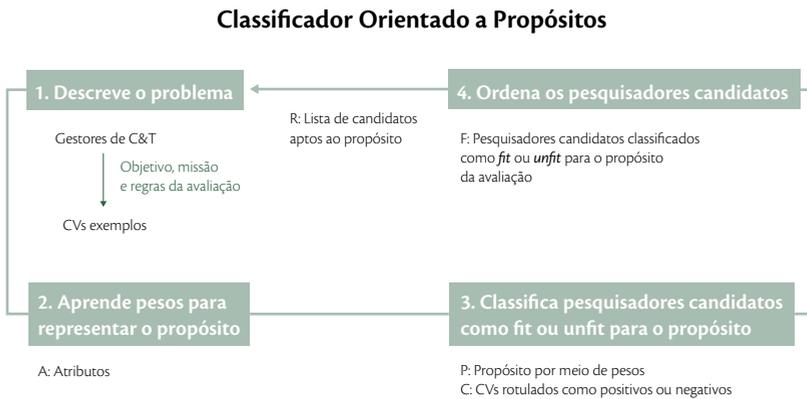
Várias aplicações baseadas em RBC têm sido desenvolvidas para diferentes áreas de conhecimento em tarefas tanto analíticas quanto sintéticas, como, por exemplo, no planejamento de processos de negócios (AGORGIANITIS *et al.*, 2016), em sistemas de recomendação (GUNAWARDENA, 2013) e em aplicações envolvendo séries temporais (GUNDERSEN, 2014). Veja também o Capítulo 4 em Richter e Weber (2013) e, para a primeira geração de aplicações, veja Von Wangenheim e Von Wangenheim (2003).

### 3. Método orientado a propósitos para mensurar a qualidade de pesquisadores – POMARQ

O método POMARQ é resultado da tese de doutorado da primeira autora deste artigo, descrito detalhadamente em Duarte (2017). A solução é implementada por meio de técnicas computacionais baseadas em aprendizagem de máquina, apoiadas em RBC. Especificamente, trata-se de uma implementação dos estágios recuperação e reutilização do Ciclo de RBC. A base da solução é um classificador, isto é, uma estrutura de dados que pode classificar um currículo de pesquisador como apto ou inapto a um específico propósito de avaliação. O classificador é composto de quatro estágios, conforme apresentado na Figura 1.

No Estágio 1, gestores de C&T estabelecem os objetivos do processo de avaliação por meio de exemplos de currículos, denominados CVs exemplos. Na perspectiva de um processo de concessão de fomento à colaboração internacional de pesquisadores para trabalharem com grupos de pesquisa em instituições de um determinado país, a solução adotada pressupõe

que, se as produções científicas dos pesquisadores destes grupos estiverem registradas em currículos, métodos de aprendizagem de máquina poderão ser aplicados para caracterizar automaticamente o contexto de tais grupos e, conseqüentemente, identificar seus requisitos de qualidade, representando o propósito da avaliação.



**Figura 1.** POMARQ – Classificador orientado a propósitos

Fonte: Adaptado de Duarte (2017).

No Estágio 2, o propósito da avaliação é representado por meio de pesos relacionados aos atributos das produções científicas dos pesquisadores. Este estágio é uma etapa de pré-processamento, que recebe como parâmetro os atributos dos CVs exemplos (*i.e.*, produções científicas registradas no currículo) e avalia a relativa relevância de cada atributo por meio de soluções de aprendizagem de pesos (ver seção 2.2). Como resultado dessa avaliação, o propósito dos CVs exemplos será refletido por meio da relevância de tais atributos.

No Estágio 3, calcula-se a medida de similaridade entre os currículos de cada candidato (CVs candidatos) e cada CV exemplo. A esta medida de similaridade chamamos de medida de fitness. Desta forma, os pesquisadores candidatos são classificados como *fit* (aptos) ou *unfit* (inaptos) ao propósito da avaliação. Ao final desse processo, é obtido um conjunto de currículos suficientemente semelhantes aos CVs exemplos.

No Estágio 4, currículos de pesquisadores candidatos, classificados como *fit*, são ordenados em uma lista classificatória. Esta lista resultante é, então, entregue aos gestores de C&T e pode ser reutilizada e adaptada em um novo ciclo de implementação do classificador.

## 4. Aplicando o método POMARQ no Currículo Lattes

Nesta seção, o método POMARQ foi aplicado sobre dados do Currículo Lattes (*lattes.cnpq.br*) e do Diretório de Grupos de Pesquisa no Brasil (DGP) (*lattes.cnpq.br/web/dgp*), de forma a simular um processo de concessão de fomento a grupos colaborativos e demonstrar a usabilidade da solução proposta.

Primeiramente, descrevem-se a fonte de dados, atributos, e os conjuntos de dados usados no cenário da aplicação. Em seguida, aplica-se o método POMARQ, que considera o propósito da avaliação e as experiências dos pesquisadores, ou seja, produções científicas mapeadas em seus currículos.

### 4.1. Dados do cenário de aplicação

A base de dados da Plataforma Lattes (*lattes.cnpq.br*) é a principal fonte de informações utilizadas no cenário de aplicação. Em julho de 2016, extraiu-se da plataforma uma amostra de 212 mil currículos de pesquisadores com doutorado concluído até o ano de 2015, além de produções científicas entre 1950 e 2015. Em seguida, foram criados dois conjuntos de dados.

O primeiro, composto de CVs exemplos, inclui currículos de 20 pesquisadores do Grupo de Pesquisa da Epidemia de Microcefalia (MERG)<sup>5</sup>, registrado no DGP<sup>6</sup> e adotado como exemplo no cenário de aplicação. O grupo MERG se tornou referência mundial em colaboração durante a epidemia de microcefalia no estado de Pernambuco, ao associar o Zica Vírus à Microcefalia (GROUP, 2016).

Para o segundo conjunto de dados, composto de CVs candidatos, foram selecionados inicialmente 15.266 currículos das cinco regiões do Brasil com as mesmas áreas de conhecimento do conjunto de CVs exemplos. Após um processo de limpeza de dados, conforme descrito em Duarte (2017), restaram 14,8 mil currículos.

Atributos referentes às produções científicas da base Lattes (*lattes.cnpq.br*) foram selecionados conforme Duarte (2017). Um subconjunto destas produções é ilustrado no Quadro 1.

5 Mais informações podem ser obtidas em: <http://www.cpqam.fiocruz.br/mergl>.

6 Mais informações podem ser obtidas em: <http://dgp.cnpq.br/dgp/espelhogrupo/2723404431935999>.

**Quadro 1.** Lista de tipos de produções científicas identificadas na base Lattes

#	Produções Científicas
1	Artigos completos publicados em periódicos
2	Trabalhos publicados em anais de eventos
3	Trabalhos não publicados em eventos
4	Livros publicados e organização de obras publicadas
5	Capítulos de livros
6	Patentes
7	Projetos de pesquisa – Auxílio financeiro
8	Projetos de pesquisa – Bolsa
9	Projetos de pesquisa – Cooperação
10	Projetos de pesquisa – Outros
11	Orientações e supervisões concluídas – Dissertação de mestrado (Orientador principal)
12	Orientações e supervisões concluídas – Dissertação de mestrado (Coorientador)
13	Orientações e supervisões concluídas – Tese de doutorado (Orientador principal)
14	Orientações e supervisões concluídas – Tese de doutorado (Coorientador)
15	Orientações e supervisões em andamento – Dissertação de mestrado (Orientador principal)
16	Orientações e supervisões em andamento – Dissertação de mestrado (Coorientador)
17	Orientações e supervisões em andamento – Tese de doutorado (Orientador principal)
18	Orientações e supervisões em andamento – Tese de doutorado (Coorientador)
19	Participações em bancas de trabalhos de conclusão – Mestrado
20	Participações em bancas de trabalhos de conclusão – Doutorado

Fonte: Adaptado de Duarte (2017).

No próximo item, são expostos os resultados da aplicação do método POMARQ sobre os dados previamente descritos.

## 4.2. Resultados da aplicação do método POMARQ

A aplicação do método POMARQ seguiu a sequência de quatro estágios especificados na descrição do método (ver Seção 3), detalhados a seguir.

No primeiro estágio foi definido o problema a ser resolvido. Especificamente, o propósito da avaliação é selecionar pesquisadores candidatos com currículos similares aos dos membros do grupo MERG (Ver seção 4.1) para trabalhar em colaboração com esses pesquisadores.

No segundo estágio, o contexto do grupo MERG foi caracterizado de forma a refletir o propósito da avaliação. Por meio de uma solução de aprendizagem de pesos (ver seção 2.2), obteve-se uma lista com a relativa relevância dos atributos dos currículos dos membros do grupo, do mais relevante ao menos relevante. A Figura 2 apresenta as 20 produções científicas mais relevantes do grupo MERG, as quais representam seu contexto de trabalho.



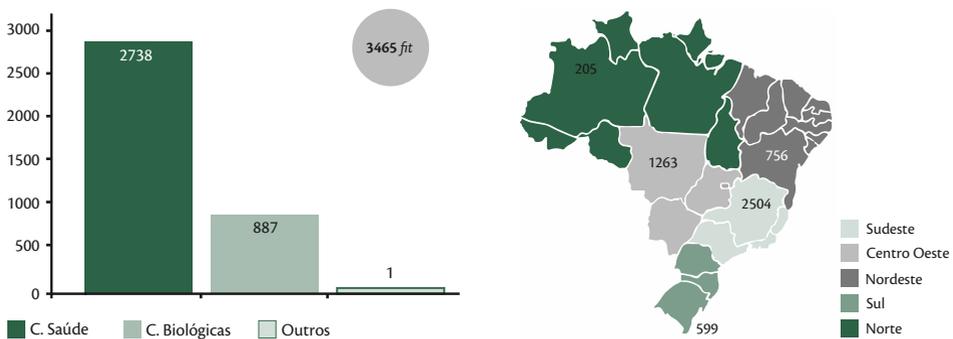
**Figura 2.** O propósito do grupo MERG representado pelo peso das 20 produções científicas mais relevantes

Fonte: Adaptado de Duarte (2017).

Ao observar as produções científicas listadas na Figura 2, é possível destacar pelo menos duas tendências. A primeira é o desenvolvimento de atividades voltadas à formação de pesquisadores, como orientações a estudantes de mestrado e doutorado. A segunda está relacionada à produção de atividades científicas, como a publicação de artigos, livros, capítulos de livros e o desenvolvimento de projetos de pesquisa.

O objetivo, no terceiro estágio, foi classificar os 14,8 mil currículos de pesquisadores candidatos como *fit* ou *unfit* ao contexto identificado no passo anterior. A medida de *fitness* foi calculada pela comparação entre os atributos do currículo de um candidato e os dos membros do grupo MERG. Esse processo resultou em 3.465 pesquisadores candidatos classificados como *fit* e 11.335 classificados como *unfit*.

A Figura 3(a) apresenta os 3.465 pesquisadores *fit* distribuídos por áreas de conhecimento e a Figura 3 (b), pelas cinco regiões do Brasil. Nesta distribuição foram considerados que um pesquisador pode atuar em mais de uma área de conhecimento, ou mesmo, estar afiliado a mais de uma instituição localizada em diferentes regiões do País. Por exemplo, um pesquisador da área de Ciências da Saúde pode também atuar na área de Ciências Biológicas.



**Figura 3.** Candidatos *fit*, aptos a trabalharem em colaboração com o grupo de pesquisa MERG, distribuídos por área de conhecimento do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e pelas cinco regiões do Brasil

Fonte: Adaptado de Duarte (2017).

O gestor do programa de fomento poderia, conforme recomendam os princípios três e seis do Manifesto de Leiden (ver Seção 1), selecionar pesquisadores de modo a atender especificidades das áreas de Ciências Biológicas e Ciências da Saúde ou mesmo características regionais no Brasil.

No quarto estágio, a lista final – ordenada de forma decrescente – de candidatos *fit* ao propósito de trabalhar em colaboração com o grupo MERG é entregue aos gestores do programa de concessão de fomento. O processo pode, então, ser encerrado, ou ter seu resultado submetido a um novo ciclo de avaliação.

## 5. Conclusões

Este artigo apresentou o método POMARQ, que é orientado ao propósito da avaliação, para mensurar a qualidade de pesquisadores. Sua usabilidade foi ilustrada por um cenário de aplicação com dados do Currículo Lattes e do Diretório dos Grupos de Pesquisa no Brasil (DGP). A solução simula um processo de concessão de fomento a pesquisadores candidatos a colaborar com o Grupo de Pesquisa da Epidemia de Microcefalia (MERG). Os resultados levam a pelo menos três significantes contribuições.

A primeira está relacionada ao uso de bases de currículos como fonte de conhecimento para avaliação da qualidade de pesquisadores. Neste sentido, o estudo abre potenciais perspectivas ao uso do Currículo Lattes.

A segunda contribuição diz respeito à capacidade da solução, conforme demonstrado, em representar os critérios da avaliação por meio da relativa relevância das produções científicas inseridas nos currículos dos colaboradores do grupo de pesquisa. Gestores de agências de fomento podem, conseqüentemente, compreender o contexto de trabalho de grupos de pesquisadores. Dessa forma, o método proposto contribui potencialmente com o DGP.

A terceira contribuição é a oferta, aos gestores de CT&I, de uma metodologia para: i) avaliar, de forma automática e por meio de técnicas de inteligência artificial aplicadas em cada estágio, uma grande quantidade de currículos; e ii) identificar os mais aptos aos critérios estabelecidos. A solução proposta poderia, por exemplo, selecionar pesquisadores para colaborar em grandes desafios da ciência.

Conforme o primeiro princípio do Manifesto de Leiden (ver Seção 1), o método proposto diz respeito a avaliações quantitativas para apoio a gestores de CT&I, assim como especialistas em avaliações qualitativas especializadas.

## Referências

AAMODT, A.; PLAZA, E. Case-based reasoning: foundational issues, methodological variations, and system approaches. **AI Communications**, v. 7, n. 1, p. 39–59, 1994.

AGORGIANITIS, I.; PETRIDIS, M., KAPETANAKIS, S.; FISH, A. Evaluating distributed methods for CBR systems for monitoring business process workflows. *In: INTERNATIONAL CONFERENCE ON CASE BASED REASONING*, 24. Atlanta Georgia, 31 oct. - 2 nov. 2016. **Proceedings...** p. 122–131, 2016.

AHA, D.W. Feature weighting for lazy learning algorithms. *In: LIU, H.; MOTODA, H. (Eds.) Feature extraction, construction and selection: a data mining perspective*. Norwell, MA: Kluwer. 1998.

AHMED, M.U.; BEGUM, S.; FUNK, P. Case studies on the clinical applications using case-based reasoning. *In: FEDERATED CONFERENCE ON COMPUTER SCIENCE AND INFORMATION SYSTEMS - FedCSIS*, 2012. **Proceedings...** IEEE, 2012. p.3–10,

BEAVER, D. DEB.; ROSEN, R. Studies in scientific collaboration: Part I. The professional origins of scientific co-authorship. **Scientometrics**, v. 1, n. 1, p. 65–84, 1978.

CARBONELL, J.G.; MICHALSKI, R.S.; MITCHELL, T.M. Machine learning: a historical and methodological analysis. **AI Magazine**, v. 4, n. 3, p. 69, 1983.

CENTRO DE GESTÃO E ESTUDOS ESTRATÉGICOS - CGEE. **Mestres e doutores 2015** - Estudos da demografia da base técnico científica brasileira. Brasília, DF: 2016.

CONFERÊNCIA INTERNACIONAL SOBRE INDICADORES DE CIÊNCIA E TECNOLOGIA, 19., STI 2016. Valência, Espanha. **Anais...** Valencia, Espanha, 14-16 set. 2016.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO - CNPq. **Painel Lattes**. Disponível em: <http://estatico.cnpq.br/painelLattes/mapa>. Acesso em: 1 jun. 2018.

DIETZ, J.S. *et al.* Using the curriculum vita to study the career paths of scientists and engineers: An exploratory assessment. **Scientometrics**, v. 49, n. 3, p. 419–442, 2000.

DUARTE, K.B. **Assessing researcher quality for collaborative purposes**. 2017. 247 f. Tese (Doutorado) – Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Florianópolis, 2017.

DUARTE, K.; WEBER, R.; PACHECO, R. C. S. Purpose-oriented metrics to assess researcher quality. *In: 21st International Conference on Science and Technology Indicators (STI2016): Peripheries, frontiers and beyond*, p.1312- 1314, 2016.

FURTADO, C.A. *et al.* A spatiotemporal analysis of Brazilian science from the perspective of researchers' career trajectories. **PLoS ONE**, v. 10, n. 10, p. 1–28, 2015.

GARFIELD, E. "Science Citation Index" - a new dimension in indexing. **Science**, v. 144, n. 3619, p. 649–654, 1964.

GARFIELD, E.; MALIN, M. Can Nobel Prize winners be predicted. *In: ANNUAL MEETING, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE*, 135. Dallas, Texas. 1968. **Proceedings...** 1968. p. 1–8.

GAUGHAN, M. Using the curriculum vitae for policy research: an evaluation of National Institutes of Health center and training support on career trajectories. **Research Evaluation**, v. 18, n. Jun., p. 117–124, 2009.

GROUP, M.E.R. Microcephaly in Infants, Pernambuco State, Brazil, 2015. **Emerging infectious diseases**, v. 22, n. 6, p. 1090–1093, 2016.

GUNAWARDENA, S. **Recommending research profiles for multidisciplinary academic collaboration**. 2013. (PhD Thesis) - Drexel University.

GUNDERSEN, O.E. **Enhancing the situation awareness of decision makers by applying case-based reasoning on streaming data**. 2014. (PhD thesis) - Norwegian University of Science and Technology.

HALL, M.A. **Correlation-based feature selection for machine learning**. 1999. (PhD thesis) - University of Waikato.

HAUSTEIN, S.; LARIVIÈRE, V. The use of bibliometrics for assessing research: Possibilities, limitations and adverse effects. *In: WELPE, I.M.; WOLLERSHEIM, J.; RINGELHAN, S.; OSTERLOH, M. (Eds.) Incentives*

**and performance: governance of research organizations.** New York: Springer International Publishing, 2015. p. 1-14.

HICKS, D. *et al.* Manifesto de Leiden sobre métricas de pesquisa. Trad. português brasileiro de Hicks *et al.* The Leiden Manifesto for research metrics. **Nature**, v. 520, p. 429-431, 2015. Disponível em: <http://www.leidenmanifesto.org/translations.html>. Acesso em: 22 set. 2018.

JURAN, J.M.; GODFREY, A.B. **Juran's quality handbook**. 5th ed., New York: McGraw-Hill, 1999.

KIRA, K.; RENDELL, L. The feature selection problem: Traditional methods and a new algorithm. **Aaai**, p. 129–134, 1992.

LANE, J. Let's make science metrics more scientific. **Nature**, v. 464, n. 7288, p. 488–9, 25 mar. 2010.

LAWANI, S.M. Some bibliometric correlates of quality in scientific research. **Scientometrics**, v. 9, n. 1-2, p. 13-25, 1986.

MÁNTARAS, L.M. *et al.* Retrieval, reuse, revision, and retention in case- based reasoning. **The Knowledge Engineering Review**. Cambridge University Press, 2005. v. 0, p. 1–24.

MOLAS-GALLART, J. Research governance and the role of evaluation: a comparative study. **American Journal of Evaluation**, p. 1–16, 2012.

PACHECO, R.C.S. **Uma metodologia de desenvolvimento de plataformas de governo para geração e divulgação de informações e de conhecimento**. Florianópolis, SC: Grupo Stela. 2003.

RICHTER, M.M.; WEBER, R.O. **Case-based reasoning: a textbook**. Berlin: Springer-Verlag, 2013.

SANDSTRÖM, U. Combining curriculum vitae and bibliometric analysis: mobility, gender and research performance. **Research Evaluation**, v. 18, n. 2, p. 135–142, 2009.

SARTORI, Rejane. **Governança em agentes de fomento dos sistemas regionais de CT&I**. 2011. 238f. Tese (Doutorado) – Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Florianópolis, 2011.

VAN NOORDEN, R. A profusion of measures. **Nature**, v. 465, n. June, p. 864–866, 2010.

VON WANGENHEIM, C.G.; VON WANGENHEIM, A. **Raciocínio baseado em casos**. Campinas, SP: Manole, 2003.

WATSON, I. Case-based reasoning is a methodology not a technology. **Knowledge-Based Systems**, v. 12, n. 5–6, p. 303–308, 1999.