

Uma Plataforma para a Extração, Integração e Análises de Dados Científicos em Acesso Aberto

Patrícia Mascarenhas Dias¹, Thiago Magela Rodrigues Dias²

¹ 0000-0002-8448-6874, Doutoranda, Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), Belo Horizonte, MG, Brasil. patriciamdias@gmail.com

² 0000-0001-5057-9936, Doutor, Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), Belo Horizonte, MG, Brasil. thiogomagela@cefetmg.br

Tipo de trabalho: comunicação

Palavras-chave: Acesso Aberto, Plataforma Lattes, Extração de Dados.

Resumo

Uma nova geração de serviços disponíveis principalmente na Web está mudando a forma de divulgar e disponibilizar a produção científica e tecnológica. Existe, atualmente, uma tendência que reforça a troca de informações e a colaboração entre as pessoas. Com a competição cada vez mais acirrada entre as instituições de pesquisa, torna-se importante para seus integrantes determinar a abrangência de suas publicações e a qualidade de seus trabalhos quando comparados aos de outros grupos, bem como identificar potenciais colaboradores com o intuito de realizar trabalhos em colaboração visando impulsionar a sua produção científica e obter melhores resultados em suas pesquisas. No entanto, o grande volume de dados sobre produção científica disponível em diferentes formatos e em diferentes repositórios dificulta a realização de estudos e consultas por parte de usuários que necessitam de uma visão unificada desses dados para, por exemplo, possibilitar a identificação de grupos de indivíduos que estejam trabalhando com determinado tema em diferentes instituições ou regiões. Além disso, o crescimento e a evolução da Web deram origem a uma grande quantidade de dados textuais pouco estruturados, heterogêneos e armazenados, sem nenhuma preocupação com padronização em diferentes repositórios. Estudos bibliométricos, principalmente em grandes repositórios bibliográficos, não são tarefas triviais tendo em vista a quantidade de dados a serem analisados e as características dos repositórios, que, em sua maioria, não possuem um padrão definido. Logo, fica evidente a dificuldade que existe para se realizar estudos abrangentes que possam apresentar de forma ampla análises sobre toda a produção científica de um grande conjunto de indivíduos que estejam vinculados a diferentes instituições ou que atuam em áreas distintas, como, por exemplo, o conjunto de todos os pesquisadores de um determinado país. Diante disso, este trabalho apresenta um estudo sobre a produção científica registrada nos currículos cadastrados na Plataforma Lattes, possibilitando, desta forma, que o framework proposto neste trabalho, possa também ser utilizado em fontes distintas de dados em acesso aberto para estudos bibliométricos e baseados em colaborações científicas.