

Uma Coleção de Dados Estratificados em Acesso Aberto para Análises Bibliométricas

Patrícia Mascarenhas Dias; patriciamdias@gmail.com; CEFET-MG
 Thiago Magela Rodrigues Dias; thiagomagela@cefetmg.br; CEFET-MG

Palavras-chave: Acesso Aberto, Produção Científica, Colaboração Científica, Plataforma Lattes.

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. Atualmente, grande parte desses estudos tem utilizado como principais fontes de dados resultados de consultas a repositórios internacionais que apresentam dados sobre trabalhos científicos, geralmente publicados em periódicos indexados. Entretanto, muitos desses repositórios negligenciam trabalhos publicados em periódicos nacionais que geralmente não são indexados e parte dos artigos publicados em anais de congressos, que constituem importante meio de publicação de algumas áreas do conhecimento.

Diante disso, este trabalho apresenta uma coleção de dados estratificados extraídos dos currículos de todos os indivíduos com doutorado concluído cadastrados na Plataforma Lattes. Essa coleção, inclui dados sobre a formação, orientações concluídas e em andamento, produção científica e colaborações desses indivíduos, possibilitando, desta forma, a realização de diversos estudos sobre esses pesquisadores brasileiros.

Para a geração da coleção de dados apresentada neste trabalho, foram coletados da Plataforma Lattes os currículos de todos os doutores ali registrados. Para a criação dos estratos de dados utilizou-se um arcabouço denominado *LattesDataXplorer* desenvolvido especificamente para a coleta, extração e tratamento de dados da Plataforma Lattes.

A Tabela 1 apresenta o quantitativo geral de todos os trabalhos registrados nos currículos dos doutores coletados para a geração da coleção.

Tabela 1 – Quantitativo dos dados dos currículos utilizados.

Tipo de Trabalho	Quantidade
Artigos em Anais de Congresso	9.051.680
Artigos em Periódico	4.660.430
Capítulos de Livro	1.054.844
Demais Trabalhos	445.894
Livros	397.934
Textos em Jornais e Revistas	858.789
Trabalhos Técnicos	1.342.416
Outras Produções Bibliográficas	625.621

Já a Tabela 2 apresenta a descrição dos estratos de dados disponibilizados. Como pode ser observado, os dados dos doutores estão organizados em conjuntos de dados que agrupam informações extraídas dos currículos e que possibilitam a realização de análises específicas. É importante ressaltar que os dados contidos nesses estratos podem ser combinados por meio dos identificadores dos respectivos currículos, propiciando um amplo escopo de análises que podem ser realizadas com os dados ora disponibilizados.

Tabela 2 – Descrição dos estratos de dados disponibilizados.

Estrato	Descrição
Formação Acadêmica	Dados sobre a formação acadêmica de cada doutor da graduação até o pós-doutorado.
Proficiência	Dados sobre nível de conhecimento (compreensão, fala, leitura e escrita) de cada doutor nos idiomas que domina.
Orientações	Dados sobre as orientações em andamento e concluídas de cada doutor nos diversos níveis de capacitação.
Produção Científica	Dados sobre a produção científica de cada doutor nos diversos tipos de veículo de publicação (periódicos, anais de conferências, livros, etc.).
Atuação Profissional	Dados sumarizados sobre a atuação profissional de cada doutor.
Colaborações	Dados sobre as colaborações científicas de cada doutor com os demais indivíduos do mesmo grupo, tendo como base as publicações produzidas no quinquênio 2012 a 2016.