

Avaliação da Gestão de Dados da Web na Análise da Produção Científica do IFNMG

Thiago de Matos Campos[1]; Abdiel Batista Santos[1]; Joice Gabrielle Souza Silva[2]; Jeancarlo Campos Leão[3]

[1]Discente. Curso Tecnólogo em Análise e Desenvolvimento de Sistemas. IFNMG campus Araçuaí; [2]Discente. Curso Técnico em Informática Integrado ao Ensino Médio. IFNMG campus Araçuaí; [3]Docente. IFNMG campus Araçuaí

Introdução

A produção científica das instituições de educação vem sendo amplamente analisada em estudos de diversas áreas do conhecimento (NEWMAN, 2001), fornecendo indicadores sobre as atividades de pesquisa e inovação dessas instituições. Constituem atividades de produção científica a publicação de: artigos em periódicos, trabalhos completos e resumos em anais de eventos, capítulos de livros, registros de propriedade intelectual, dentre outros.

Felizmente, grande parte dos dados da produção científica brasileira estão publicados na Web, em especial, na plataforma Lattes[†] que é responsável por armazenar informações curriculares da comunidade científica nacional. Atualmente, essa é uma das mais importantes fontes para coleta de dados sobre a produção científica nacional e do Instituto Federal do Norte de Minas Gerais (IFNMG).

Contudo, o processo manual de obtenção dos dados de produção científica dos campus do IFNMG a partir do Currículo Lattes não é eficiente, ainda que sejam estruturados e armazenados em bancos de dados. A cada nova demanda por essas informações por diferentes setores e comissões dos IFNMG é necessário coletar manualmente e exaustivamente os dados através de uma busca por todos os currículos de interesse, além da necessidade de ler cada currículo para identificar e selecionar informações específicas. Em seguida, essas informações ainda precisam ser tratadas para evitar duplicações, ambiguidades, erros e dados faltantes. Neste contexto, surge a nossa questão de pesquisa: como a gestão de dados pode ajudar na análise da produção científica dos campi do IFNMG?

Diversos modelos, linguagens e técnicas permitem lidar com dados em um ambiente heterogêneo e distribuído como a Web e esses constituem os principais mecanismos e fundamentos por trás da Gerência de Dados na Web (GDW) (ABITEBOUL et al., 2011). Assim, nossa principal hipótese é que a GDW pode facilitar o processo de análise da produção científica desde seu início e em todo o seu ciclo de vida (BARBIERI, 2013) apoiada por esses mecanismos.

Quantificar a produção científica permite estimar a pesquisa no campus e fornecer subsídios para tomada de decisão. Por exemplo, indicadores de produção científica são importantes na avaliação de cursos superiores, na revalidação de programas de pós-graduação, para participar de diversos editais de fomento à pesquisa no Brasil (DIAS, 2017) ou a outras demandas relacionadas, locais aos campus do IFNMG como as da coordenação de pesquisa. Além disso, por conter a trajetória acadêmica do pesquisador, permite a análise do seu mérito e competência que são considerados em editais de afastamento para capacitação, processos de progressão e financiamento de bolsas e projetos.

Assim, o objetivo deste trabalho é avaliar uma proposta de gestão de dados sobre a produção científica do IFNMG a partir da extração dos currículos da plataforma Lattes. Mais especificamente, objetivamos desenvolver e avaliar um processo que permite sumarizar de forma prática e rápida o perfil de produção acadêmica de cada campus do IFNMG.

---

*Apresentado por: Thiago de Matos Campos

†A Plataforma Lattes do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) mantém o Currículo Lattes que se tornou em agosto de 1999 o currículo oficial do CNPq.



Metodologia

Para a gestão de dados foi utilizada como principal ferramenta a plataforma de Gestão de Dados Científicos - GeDai. A plataforma GeDai permite automatizar a extração dos elementos do currículo Lattes e emitir relatórios sobre um conjunto de pesquisadores. Assim, com apoio da plataforma GeDai, o seguinte processo foi seguido:

Primeiro, uma extração inicial e única foi feita para obter automaticamente os currículos identificados para o campus escolhido, que nesse caso foi o campus Araçuaí. Em seguida, a partir do relatório de currículos Lattes dos servidores do campus, comparamos a lista de de nomes nos currículos atualmente coletados com a lista de servidores disponíveis na página Web do campus. Essa etapa permite identificar currículos que foram localizados pela plataforma em razão de algum dado faltante no campo de afiliação do servidor. Então, registramos manualmente o identificador do currículo não encontrado na plataforma e o associamos ao campus Araçuaí. Solicitamos a coleta novamente ao GeDai. Por fim, com o relatório anual de produção científica emitido pelo GeDai, obtivemos a lista com todos os currículos associados ao campus e a contagem de produção científica agregada por tipo de produção como artigos, livros, etc.

A avaliação da qualidade do processo proposto consiste em duas abordagens. Na primeira, foi feita inspeção individual de uma amostra de currículos, objetivando estimar a capacidade da ferramenta GeDai para extrair corretamente os dados contidos nos currículos. Na segunda abordagem, através de entrevista com usuários do GeDai, objetivamos conhecer a satisfação do seu usuário quanto às características de usabilidade da plataforma e a eficácia do processo. Inicialmente foram elaboradas duas entrevistas.

Resultados Parciais

Para as análises da produção científica do IFNMG, considerou-se inicialmente o conjunto de currículos dos servidores do campus Araçuaí cadastrados na Plataforma Lattes. Com isso, ressalta-se que o conjunto de currículos analisados neste trabalho compreende a grande parte dos servidores em atuação no campus Araçuaí.

Na avaliação de usabilidade, os usuários do sistema GeDai consideram sua interface amigável e de fácil uso. O processo se apresentou eficaz para obter as informações dos currículos que estavam atualizados e tornou mais eficiente a obtenção de dados de produção científica.

Contudo, foram identificados currículos que não foram atualizados recentemente, o que implica em um desafio para o melhor aproveitamento do processo proposto. Primeiro, porque dificulta a identificação do campus de lotação atual do servidor. Segundo, um currículo desatualizado impede que o processo seja feito com maior proveito da automação. Isso demonstra a necessidade de incentivo à manutenção de dados atualizados nos currículos da Plataforma Lattes.



Considerações Finais

Diante das dificuldades para mensurar rapidamente a produção científica do IFNMG, um processo automatizado pode contribuir para a obtenção de subsídios que fortaleçam as atividades de pesquisa em cada campus. Foi proposto um processo de extração automatizada, tratamento, seleção e apresentação de dados da produção científica do IFNMG a partir dos dados extraídos dos currículos da plataforma Lattes utilizando a ferramenta GeDai. Assim, o objetivo deste trabalho é avaliar a aplicação do processo proposto para gestão de dados de caracterização da produção científica do IFNMG.

O processo proposto foi avaliado sob o aspecto da facilidade de uso e eficácia dos resultados produzidos, representados principalmente por um relatório de sumarização da produção científica, agregado por tipo de produção. Além disso, foram feitas entrevistas com usuários do sistema o que permitiu verificar a satisfação sobre a qualidade dos dados obtidos a partir desse processo. Além disso, os currículos desatualizados se apresentaram como obstáculo ao aproveitamento total do potencial do processo.

Na continuidade deste trabalho, pretendemos avaliar a aplicação do processo de gestão de dados em outros campus do IFNMG. Além disso, será considerado um número maior de entrevistados para validar a satisfação em relação aos resultados produzidos.

Referências

ABITEBOUL, S. et al. *Web data management*. [S.l.]: Cambridge University Press, 2011.

BARBIERI, C. Uma visão sintética e comentada do data management body of knowledge (dmbok). 2013.

NEWMAN, M. E. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 98, n. 2, p. 404–409, 2001.