



Evento	Salão UFRGS 2022: SIC - XXXIV SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2022
Local	Campus Centro - UFRGS
Título	Análise da pesquisa Brasileira via plataforma Lattes: web scraping e banco de dados semi-estruturados
Autor	LUIZ ALMIR ZANELLA DE PAULA
Orientador	TAIANE SCHAEGLER PRASS

O objetivo da pesquisa é a elaboração de rankings para a análise da qualidade das pesquisas acadêmicas das universidades brasileiras. O motivo desta é melhor conhecermos a realidade da academia nacional e as diferentes realidades enfrentadas por pesquisadores. Para isto, utilizamos variadas técnicas. Em primeiro lugar, efetuamos a coleta de dados da plataforma Lattes utilizando técnicas de web scrapping, e depois, buscamos informações relevantes para futuras análises, por exemplo, ano de publicação, jornal e número de autores por publicação, bem como a instituição à qual o pesquisador está vinculado (isto tudo através de técnica conhecida como text mining) para cruzar essas informações acerca dos artigos publicados com os do site Scimago (site que atribui uma nota para as mais diversas revistas científicas mundo afora). Em um segundo momento, para a organização, limpeza e análise dos dados, aprendemos e utilizamos o MongoDB, software de banco de dados orientado a documentos, através do software R com o pacote "Mongolite". Após uma limpeza preliminar, efetuamos a análise dos dados coletados e passamos a elaborar diversos rankings, parciais frise-se, sobre a qualidade da produção de artigos científicos no país, lembrando que neste trabalho, o foco é apenas em doutores vinculados a instituições de ensino reconhecidas pelo Ministério da Educação. Assim, neste trabalho iremos apresentar um resumo de tudo que foi estudado e elaborado até o momento.