



XXXIII SIC SALÃO INICIAÇÃO CIENTÍFICA

Evento	Salão UFRGS 2021: SIC - XXXIII SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2021
Local	Virtual
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

Universidade Federal do Rio Grande do Sul - UFRGS

Autor: Luiz Almir Zanella de Paula

Orientadora: Taiane Schaedler Prass

Título: Análise da Pesquisa Brasileira via Plataforma Lattes: Web Scraping e Banco de Dados semi-estruturados.

Resumo: o objetivo da pesquisa é a elaboração de rankings para a análise da qualidade das pesquisas acadêmicas das universidades brasileiras. 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 (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 estamos utilizando o MongoDB, software de banco de dados orientado a documentos, através do software R com o pacote Mongolite. Neste trabalho iremos apresentar um resumo de tudo que foi estudado e elaborado até o momento.