

WHILEDGP, UTILIZAÇÃO DA TÉCNICA DE SCRAPING E AUTOMAÇÃO NA COLETA DE DADOS NO SITE DO CNPQ

Lucas Da Silva Maciel¹
Renato Farias De Paiva²
Antonio Paulo Uamba³
Pedro Bruno Silva Lemos⁴
Allberson Bruno De Oliveira Dantas⁵

RESUMO

Percebendo a necessidade de avaliar a produção científica gerada a partir dos grupos de pesquisa da Universidade da Integração Internacional da Lusofonia Afro-brasileira (UNILAB), registrados no Diretório de Grupos de Pesquisa (DGP/CNPq), foi desenvolvido o projeto intitulado como "WhileDGP". Este, por sua vez, destina-se a coletar, de forma automatizada, informações dos grupos, bem como, dados básicos de todos os pesquisadores registrados naquele grupo e suas produções científicas, declaradas na plataforma Lattes. O projeto foi desenvolvido em Python. No projeto, foram utilizadas técnicas de Business Intelligence (Inteligência de Negócio), Mineração de Dados e Visualização de Dados. A mineração dos dados na web foi feita através de Scraping. Nos últimos testes realizados, foram recuperados dados de todos os 128 grupos da UNILAB, registrando 756 pesquisadores de uma maneira totalmente automatizada e rápida. Nos dados obtidos dos grupos constam, dentre outras informações: o título do grupo, situação, líder(es), área predominante. Já nos dados dos pesquisadores são recuperados: link para o currículo lattes, link para o espelho do pesquisador, onde é possível acessar os dados resumidos da produção de cada pesquisador, dentre outros. A partir dos dados coletados, foi possível ter uma visão holística de toda a produção científica da Unilab, figurando assim como grande aliado à Gestão Institucional, sobretudo nas iniciativas desenvolvidas pela Coordenação de Pesquisa da PROPPG/Unilab. Como trabalho futuro, serão desenvolvidas análises de Aprendizagem de Máquina visando traçar fielmente os perfis dos pesquisadores da Unilab, bem como predizer o panorama de publicações e suas áreas para os próximos anos.

Palavras-chave: WhileDGP Bot Automação Scraping Python .

Universidade da Integração Internacional da Lusofonia Afro-brasileira (UNILAB), Instituto de Engenharias e Desenvolvimento Sustentável (IEDS), Discente, lucasmaciel6690@gmail.com¹
Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB), Instituto de Engenharias e Desenvolvimento Sustentável (IEDS), TAE, renatopaivabv@gmail.com²
Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB), Instituto de Engenharias e Desenvolvimento Sustentável (IEDS), Discente, martinsuanba@gmail.com³
Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB), Instituto de Engenharias e Desenvolvimento Sustentável (IEDS), TAE, pedrobruno_lemos@yahoo.com.br⁴
Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB), Instituto de Engenharias e Desenvolvimento Sustentável (IEDS), Docente, allberson@unilab.edu.br⁵

