

Tema de Trabalho Prático - Mineração de Dados

Guilherme Aguilar de Oliveira

Luiz Guilherme

Maycom da Luz de Deus

(1) Qual é o problema?

A cada ano, são publicadas diversas monografias da UTFPR. É natural que os temas variem por ano por causa de diversos fatores internos e externos ao ambiente acadêmico. Por exemplo, hipotetizamos que o tema de metodologias ágeis terá uma relevância maior nos últimos 10 anos ou o tema de LLM's terá relevância maior há 5 anos. Sabendo disso, queremos entender os temas/assuntos de pesquisa mais relevantes por ano.

(2) Qual é a fonte de dados que deseja estudar e quais dados estão disponíveis?

Como pretende

obter os dados?

A fonte de dados é o Repositório Institucional da Universidade Tecnológica Federal do Paraná (RIUT). <https://repositorio.utfpr.edu.br/jspui/>

(3) O que você pretende estudar na base de dados escolhida?

Estudar os temas/assuntos de pesquisa mais relevantes por ano.

(4) Quais técnicas de mineração de dados pretende aplicar?

Para coletar os dados, faremos web scraping para coletar os metadados dos documentos como título, resumo, autor e baixar as monografias. Após isso, será feita uma limpeza do texto dos resumos e dos pdfs. Com bastante texto, faremos uma extração de tópicos com o LDA e BERTopics. Com os tópicos e metadados em mãos, faremos uma análise exploratória dos dados tentando encontrar padrões interessantes em como os tópicos evoluem ao longo dos anos e décadas separados por curso.