



Universidade Federal do Ceará  
Centro de Ciências  
Departamento de Computação  
Bacharelado em Engenharia de Computação

## **CK0223 – Mineração de Dados 2021.1**

### **Trabalho 1: Extração de Dados**

#### **1. Descrição:**

Este trabalho consiste na construção de um dataset, a partir de uma ou mais fontes de dados (que podem ser estruturadas ou não).

#### **2. Sugestões de Datasets**

1. Para os alunos de pós-graduação que já possuem um tema definido, aconselha-se a construção de um dataset relacionado ao seu tema de pesquisa.
2. Os demais estudantes deverão construir datasets relacionados a área da Cultura. Neste sentido, algumas alternativas são:
  - a. Dataset contendo textos (e metadados) acerca da Lei de Incentivo à Cultura, conhecida como Lei Rouanet (Lei nº 8.313 do dia 23 de dezembro de 1991). Esses textos podem ser obtidos a partir do Twitter, Facebook, Instagram, WhatsApp, Blogs e/ou sites de notícias (incluindo a mídia oficial e não oficial).
  - b. Dataset contendo textos (e metadados) acerca da Lei Aldir Blanc (LEI nº 14.017, DE 29 DE JUNHO DE 2020). Esses textos podem ser obtidos a partir do Twitter, Facebook, Instagram, WhatsApp, Blogs e/ou sites de notícias (incluindo a mídia oficial e não oficial).
  - c. Dataset contendo dados dos editais da Secretaria de Cultura do Estado do Ceará (SECULT-CE), incluindo dados dos projetos selecionados.
  - d. Dataset contendo dados dos editais da Secretaria de Cultura do Município de Fortaleza (SECULT-FOR), incluindo dados dos projetos selecionados.
  - e. Dataset contendo dados da Plataforma Mapa Cultural do Ceará.
  - f. Dataset contendo dados de pontos de cultura, agentes de cultura, projetos culturais, etc. Esses textos podem ser obtidos a partir do Twitter, Facebook, Instagram, WhatsApp, Blogs e/ou sites de notícias (incluindo a mídia oficial e não oficial).

PS: Outros temas podem ser sugeridos, mas a aceitação fica a cargo do professor da disciplina.

## 2. Avaliação

Espera-se com a realização deste trabalho que cada estudante elabore e entregue (de forma digital) os seguintes documentos:

- Arquivo CSV contendo os dados do Dataset;
- Jupyter Notebook contendo os Scripts Python utilizados para obter os dados.
- Apresentação baseada em slides descrevendo as atividades desenvolvidas.

A avaliação deste trabalho se dará em três etapas:

1ª. Apresentação Oral do Dataset: Cada estudante disporá de 10 minutos para apresentar um resumo do dataset gerado e processo utilizado. O estudante pode utilizar *slides* e notebooks.

2ª. Avaliação do Dataset: O professor da disciplina irá avaliar a qualidade do dataset gerado pelo estudante, bem como dos códigos implementados.

A avaliação do trabalho irá envolver os seguintes quesitos:

- Abrangência e Organização do Dataset
- Qualidade dos Códigos Utilizados
- Clareza do Texto (Tutorial) Utilizado para Descrever o Dataset e seu Processo de Construção
- Domínio do Tema

## 3. Data da Apresentação: 28/06/2021

- PS. Não serão aceitos trabalhos que não forem apresentados.
- PS. Cada estudante será responsável pela disponibilização do ambiente (software e hardware) necessário para a apresentação do trabalho.

“A Educação, qualquer que seja ela, é sempre uma teoria do conhecimento posta em prática”.

**Paulo Freire**

