INE5454 – TÓPICOS ESPECIAIS EM GERÊNCIA DE DADOS

CARINA DORNELES

CARINA.DORNELES@UFSC.BR

INE5454 – Tópicos Especiais em Gerência de Dados

OUTLINE

- Plano de Ensino
  - Objetivos
  - Programa
  - Metodologia
  - Avaliação

OBJETIVOS

- Geral
  - Apresentar aos alunos uma visão teórica e prática do processo de Web Scraping, detalhando os passos de coleta de dados da Web, tratamento posterior destes dados

- Específicos
  - Apresentar os passos envolvidos em um processo de Web Scraping
  - Possibilitar o desenvolvimento de um processo completo de Web Scraping

PROGRAMA

1. Web Scraping
    - Web Crawling
    - Extração de Dados
    - Data Curation
    - Formas de uso
3. Desenvolvimento do processo de Web Scraping
    - Definição do projeto
    - Desenvolvimento da infraestrutura de implementação
    - Implementação das técnicas de Web Scraping
    - Testes de execução

METODOLOGIA

- Aulas síncronas e assíncronas
  - Aulas síncronas: ao vivo na sala criada na RNP
    - https://conferenciaweb.rnp.br/webconf/ine5454-06208-topicos-especiais-em-gerencia-de-dados
  - Aulas assíncronas:
    - Vídeo aulas
    - Textos informativos + questionário de atividades

CRONOGRAMA

- 1ª a 4ª semana
  - aulas expositivas síncronas ou assíncronas;
- 5ª a 15ª semana
  - reuniões de acompanhamento do desenvolvimento dos projetos;
  - horários específico por dupla, ou indivíduo, de trabalho (30 minutos por dupla/indivíduo);
- 16ª a 18ª semana
  - envio dos trabalhos, avaliação do professor e reuniões de feedback.

AVALIAÇÃO

- A cada reunião:
  - apresentação/discussão do andamento do trabalho
  - cada integrante do grupo deve explanar o que foi produzido e/ou pode ser questionado pelo professor
- Desempenho individual nas reuniões de desenvolvimento e apresentação do trabalho
  - Participação ativa em todos os encontros e na apresentação: nota 10
  - Ausência ou falta de participação a cada 2 encontros ou na apresentação final: perda de até um (1) ponto na nota
- Trabalho a ser desenvolvido:
  - grupos de 2 alunos, ou individual
  - Apresentação: gravação de um vídeo de 10 a 15 min. explicando o desenvolvimento (será fornecido modelo de apresentação)

REFERÊNCIAS BIBLIOGRÁFICAS

1. Ricardo Baeza-Yates, Berthier Ribeiro-Neto. Modern Information Retrieval. 1ª Edição – 1999.
2. Castilho, Carlos. Web Crawling. IN: Ricardo Baeza-Yates, Ricardo; Ribeiro-Neto, Berthier. Modern Information Retrieval. Chapter 2, 2010.
3. Robert Baumgartner and Wolfgang Gatterbauer and Georg Gottlob. Data Extraction System. Encyclopedia of Database Systems. 2009.
4. Freitas A., Curry E. (2016) Big Data Curation. In: Cavanillas J., Curry E., Wahlster W. (eds) New Horizons for a Data-Driven Economy. Springer, Cham