



PLANO DE ENSINO – PPGICS – DISCIPLINAS

() VERÃO (X) INVERNO () ELETIVA () OBRIGATÓRIA

IDENTIFICAÇÃO - REMOTA			
Disciplina: Mineração de dados: coleta e mineração de dados para composição de infraestrutura da informação			
Código: PGICS-DM028	Créditos: 2	Carga Horária: 30	Período Início: 10/07/2023 Término: 21/07/2023 Dia da Semana: Segunda-feira a Sexta-feira Horário: 9h as 12h
Código: ICS-DM154	Créditos: 2	Carga Horária: 60	
Coordenadores da Disciplina: Maria Cristina Soares Guimrães Professor: Janio Gustavo Barbosa			
Curso: () Mestrado () Doutorado		Núcleo Comum (X)	
Linha 1 (X)		Linha 2 () Linha 3 ()	

EMENTA DA DISCIPLINA (PREENCHER SOMENTE SE FOR ELETIVA REGULAR)

--

EMENTA DO CURSO

Caracterização das soluções de automação de processos para a saúde; Estudos de bases científicas estruturadas: coleta em bases científicas; Limpeza de dados; Introdução a mineração de dados – conceitos básicos; Definição de bases estruturadas, semiestruturadas e não estruturadas. Etapas de descoberta do texto; Recuperação de texto; Visão geral sobre mineração de dados textuais. Técnicas de pré-processamento de textos; Introdução a aplicação da mineração de texto: Análise de documentos jurídicos: diários de justiça e diários oficiais da união. Identificando novas fontes de informação para a saúde e captura de dimensões diversas na produção de conhecimento no campo da saúde. Infraestrutura da informação e elementos de alta disponibilidade, escalabilidade e reprodutibilidade. Características de uma infraestrutura da informação e desenvolvimento de bases de dados abertos. Sistemas copiloto e computação cognitiva: GPT, Microsoft copilot. Sistemas legais em marcos regulatórios de mineração de dados – LGPD e RGPD.



OBJETIVOS

- Explorar e desenvolver sistemas que possam extrair informações na área da saúde.
- Compreender as particularidades de uma base estruturada e uma base não estruturada
- Saber reconhecer, coletar, limpar e qualificar dados de bases estruturadas.
- Discutir questões relativas à coleta e tratamento de dados em bases estruturadas
- Compreender o processo de automatização do processo de descoberta de conhecimento em texto.
- Reconhecer o potencial do text mining para área da saúde e suas diversas aplicações.
- Compreender aspectos teóricos e metodológicos de uma infraestrutura da informação.
- Compreender e discutir aspectos da informação científica em sistemas cognitivos automatizados.

BIBLIOGRAFIA BÁSICA

BARCELOS, Paulo Cesar A. Metodologia ou tecnologia?. *In: CONGRESSO INTERNACIONAL ABED DE EDUCAÇÃO A DISTÂNCIA*, 12., 2005, Florianópolis. **Anais [...]**. Florianópolis: ABED, 2005.

BERRY, Michael W. (ed.). **Survey of Text Mining: Clustering, Classification, and Retrieval**. New York, NY: Springer, 2003.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. **Data Mining: um guia prático**. Rio de Janeiro: Editora Campus; [S. l.]: Elsevier, 2005.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data Mining: Concepts and Techniques**. Waltham: Elsevier, 2006.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data Mining: Concepts and Techniques**. Waltham: Elsevier, 2012.

PORTER, Alan L.; CUNNINGHAM, Scott W. **Tech mining: exploiting new technologies for competitive advantage**. Hoboken, NJ: John Wiley&Sons, 2004.

PORTER, Alan L. *et al.* Minería tecnológica: maneras múltiples para explotar los recursos de la Ciencia, Tecnología e Información. *In: MEDINA VÁSQUEZ, Javier; SÁNCHEZ TORRES, Marcela.* (org.). **Sinergia entre la prospectiva tecnológica y la vigilancia tecnológica e inteligencia competitiva**. Bogotá: Colciencias, 2008. p. 223-254.

RINO, Lucia Helena M. Sumarização automática de textos em português. *In: Anais do II ENCONTRO PARA O PROCESSAMENTO COMPUTACIONAL DO PORTUGUÊS ESCRITO E FALADO*, 2., 1996, Curitiba. **Anais [...]**. Curitiba: Cefet PR, 1996. p.109-119.

STAR, Susan L.; LAMPLAND, Martha. Reckoning with standards. *In: LAMPLAND, Martha; STAR, Susan L.* (ed.). **Standards and their stories: how quantifying, classifying, and formalizing practices shape everyday life**. Ithaca: Cornell University Press, 2009. p. 3-25.

STAR, Susan L.; RUHLER, Karen. Steps toward an ecology of infrastructure: design and access for large information spaces. **Information Systems Research**, [S. l.], v. 7, n. 1, p. 111-34, 1996. Disponível em: <https://pubsonline.informs.org/doi/abs/10.1287/isre.7.1.111>. Acesso em: 4 ago. 2022.



BIBLIOGRAFIA COMPLEMENTAR (opcional)

IRAMINA, Aline. RGPD v. LGPD: Adoção estratégica da abordagem Responsiva na elaboração da Lei Geral de Proteção de Dados do Brasil e do Regulamento Geral de Proteção de Dados da União Europeia. **Revista de Direito, Estado e Telecomunicações**, Brasília, v. 12, n. 2, p. 91-117, out. 2020.

MOROZOV, Evgeny. **Big Tech**: a ascensão dos dados e a morte da política. São Paulo: Ubu, 2018. 189 p. ISBN 978 85 7126 012 2.

EMPOLI, Giuliano da. **Os engenheiros do caos**. São Paulo: Vestígio, 2002.

CRITÉRIOS DE AVALIAÇÃO

O trabalho de conclusão consiste na elaboração de um texto descritivo das etapas de trabalho em forma de projeto de intervenção de livre escolha, para resolução de problema por meio de Data Mining ou Text Mining em plataforma colaborativa on-line.

CRONOGRAMA DAS AULAS REMOTA

1ª Aula 10/07	Introdução a disciplina, o que é dado, comportamento, apropriação e compartilhamento. a. Bases de dados estruturadas / exemplos de bancos de dados / exemplos de modelo relacional b. Bases de dados não estruturadas - semântica / morfologia / afetividade (particularidade) c. Bases de desenvolvimento de uma Infraestrutura da informação: Objetos infraestruturante e bases de dados abertas d. LGPD e a discussão sobre bancos de dados coletados.
2ª Aula 11/07	INFRAESTRUTURA DE INFORMAÇÃO ASPECTOS TEÓRICOS a. A necessidade de infraestruturas de informação b. Exemplos e desenvolvimento de II integrada c. Bases de dados abertas e o conceito de dados compartilhados d. Desenvolvimento e implementação de II e. Uso dos dados pelo Estado e a catástrofe informacional
3ª Aula 12/07	MINERAÇÃO DE TEXTOS: CONCEITO, PROCESSO E APLICAÇÕES (teoria) 1. O que é 2. Exemplos 3. Particularidade 4. Possibilidades 5. <i>Clusterização</i> 6. Análise dos agrupamentos / categorização 7. Aplicações (prática em linguagem R) : análise exploratória dos dados.



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz



4ª Aula 13/07	TÉCNICAS UTILIZADAS (teoria / prática) 3.1 pré-processamento: 3.2 tokenização 3.3 stemming 3.4 stopwords 3.5 BoW - Bag of Words 3.6 processamento
5ª Aula 14/07	CAPTAÇÃO DE DADOS: CRAWLING / WEB SCRAPING (teoria / demonstrativa)
6ª Aula 17/07	MINERANDO TEXTO EM R: APLICAÇÕES (teoria / prática)
7ª Aula 18/07	MINERANDO TEXTO EM R: APLICAÇÕES (teoria / prática) Formando banco de dados e categorizando os dados
8ª Aula 19/07	Formando banco de dados e categorizando os dados
9ª Aula 20/07	PROCESSO DE MINERAÇÃO DE TEXTO (prática)
10ª Aula 21/07	PROCESSO DE MINERAÇÃO DE TEXTO (prática) ETAPAS DA MINERAÇÃO DE TEXTO (teoria / prática)

Rio de Janeiro, 22 de maio de 2023.