

PERÍODO: 2020.2		
CENTRO UNIVERSITÁRIO CTC DEPARTAMENTO DE INFORMÁTICA	2020.2	
INF2390	Tópicos em Bancos de Dados IV – Tópicos em “Big Data” Prof. Marco Antonio Casanova, Antonio Luz Furtado	
	CARGA HORÁRIA TOTAL: 45	CRÉDITOS: 3
	Título do Tópico: Mineração de Grandes Conjuntos de Dados Pré-requisito:	
OBJETIVOS	Esta disciplina aborda tecnologias para minerar grandes volumes de dados.	
EMENTA	Mineração de dados; Algoritmos distribuídos para grandes volumes de dados; Pesquisa por similaridade; Processamento de sequencias de dados; Motores de busca; Mineração de conjuntos frequentes de itens de dados; Agrupamento de grandes volumes de dados multidimensionais; Sistemas de recomendação; Análise e mineração de padrões estruturais em grandes grafos.	
PROGRAMA	<ul style="list-style-type: none"> • O que é mineração de dados • Sistemas de arquivos distribuídos • Algoritmos por <i>map-reduce</i> para grandes volumes de dados • Pesquisa por similaridade: métricas de similaridade, <i>minhashing</i> e <i>locality sensitive hashing</i>; <i>KMV-Synopses</i> • Processamento de sequencias de dados • A tecnologia de motores de busca • O algoritmo de <i>PageRank</i> do Google • Mineração de conjuntos frequentes de itens de dados: regras de associação, cestas de produtos e o algoritmo <i>A-Priori</i> e suas variações • Algoritmos para agrupamento de grandes volumes de dados multidimensionais • Sistemas de recomendação • Algoritmos para analisar e minerar redes sociais 	

AVALIAÇÃO	O desempenho do aluno será avaliado por seminários e uma monografia preparada para o curso.
BIBLIOGRAFIA PRINCIPAL	<p>Rajaraman, A.; Leskovec, J.; Ullman, J.D. Mining of Massive Datasets. Cambridge University Press (2014). ISBN: 9781107015357 (disponível em http://www.mmms.org)</p> <p>Porto, F. Algoritmos e Modelos de Programação em Big Data. In: Jornadas de Atualização em Informática 2017, Porto Alegre/RS, Sociedade Brasileira de Computação – SBC. ISBN: 978-85-7669-374-1</p>
BIBLIOGRAFIA COMPLEMENTAR	<p>Hopcroft, J., Kannan, R. <i>Foundations of Data Science</i>. Version 21/8/2014. (disponível em research.microsoft.com/en-US/people/kannan/book-no-solutions-aug-21-2014.pdf)</p> <p>Han, J. e Kamber, M. <i>Data Mining: Concepts and Techniques</i>, 3rd ed. The Morgan Kaufmann Series in Data Management Systems (2011). ISBN 978-0123814791 (disponível em http://web.engr.illinois.edu/~hanj/bk3/)</p> <p>Karau, H., Konwinski, A., Wendell, P., Zaharia, M. <i>Learning Spark</i>. O'Reilly (2015). ISBN-13: 978-1449358624</p> <p>Apache Spark: Lightning-fast cluster computing (disponível em https://spark.apache.org)</p>