

|   |   |   |
|---|---|---|
| PERÍODO: 2019.2   |   |  |
| CENTRO UNIVERSITÁRIO CTC<br>DEPARTAMENTO DE INFORMÁTICA |   |   |
| <b>INF2390</b>  | Tópicos em Bancos de Dados IV – Tópicos em “Big Data”<br>Prof. Marco Antonio Casanova, Antonio Luz Furtado  |   |
|   | CARGA HORÁRIA TOTAL: 45   | CRÉDITOS: 3   |
|   | Título do Tópico: Mineração de Grandes Conjuntos de Dados<br>Pré-requisito:   |   |
| <b>OBJETIVOS</b>  | Esta disciplina aborda tecnologias para minerar grandes volumes de dados.   |   |
| <b>EMENTA</b>   | 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.   |   |
| <b>PROGRAMA</b>   | <ul style="list-style-type: none"> <li>• O que é mineração de dados</li> <li>• Sistemas de arquivos distribuídos</li> <li>• Algoritmos por <i>map-reduce</i> para grandes volumes de dados</li> <li>• Pesquisa por similaridade: métricas de similaridade, <i>minhashing</i> e <i>locality sensitive hashing</i>; <i>KMV-Synopses</i></li> <li>• Processamento de sequencias de dados</li> <li>• A tecnologia de motores de busca</li> <li>• O algoritmo de <i>PageRank</i> do Google</li> <li>• Mineração de conjuntos frequentes de itens de dados: regras de associação, cestas de produtos e o algoritmo <i>A-Priori</i> e suas variações</li> <li>• Algoritmos para agrupamento de grandes volumes de dados multidimensionais</li> <li>• Sistemas de recomendação</li> <li>• Algoritmos para analisar e minerar redes sociais</li> </ul> |   |

|                                  |   |
|----------------------------------|---|
| <b>AVALIAÇÃO</b>                 | O desempenho do aluno será avaliado por seminários apresentados em sala e uma monografia preparada para o curso.  |
| <b>BIBLIOGRAFIA PRINCIPAL</b>    | <p>Rajaraman, A.; Leskovec, J.; Ullman, J.D. Mining of Massive Datasets. Cambridge University Press (2014). ISBN: 9781107015357 (disponível em <a href="http://www.mmms.org">http://www.mmms.org</a>)</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>   |
| <b>BIBLIOGRAFIA COMPLEMENTAR</b> | <p>Hopcroft, J., Kannan, R. <i>Foundations of Data Science</i>. Version 21/8/2014. (disponível em <a href="http://research.microsoft.com/en-US/people/kannan/book-no-solutions-aug-21-2014.pdf">research.microsoft.com/en-US/people/kannan/book-no-solutions-aug-21-2014.pdf</a>)</p> <p>Han, J. e Kamber, M. Data Mining: Concepts and Techniques, 3rd ed. The Morgan Kaufmann Series in Data Management Systems (2011). ISBN 978-0123814791 (disponível em <a href="http://web.engr.illinois.edu/~hanj/bk3/">http://web.engr.illinois.edu/~hanj/bk3/</a>)</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 <a href="https://spark.apache.org">https://spark.apache.org</a>)</p> |