MINISTÉRIO DA EDUCAÇÃO

UNIVERSIDADE FEDERAL DO PARANÁ

SETOR DE CIÊNCIAS EXATAS

Departamento de Estatística

Ficha 1 (permanente)

Disciplina:**Mineração de texto**							Código: CE040
Natureza: ( ) Obrigatória (X) Optativa		(X) Semestral ( ) Anual ( ) Modular					
Pré-requisito: Nenhum	Co-requisito: Nenhum	Modalidade: (X) Presencial ( ) Totalmente EAD ( ) CH em EAD: _____					
CH Total: 60 CH Semanal: 04	Padrão (PD): 60	Laboratório (LB):	Campo (CP):	Estágio (ES):	Orientada (OR):	Prática Específica (PE):	Estágio de Formação Pedagógica (EFP):

EMENTA

Motivação, história e tendências. Medidas descritivas para texto. Noções de linguística computacional. Abordagens para a mineração de texto: Bag of words (BOW) e NLP. Manipulação de cadeias de caracteres. Preprocessamento para BOW. Visualização em mineração de texto. Análise de sentimentos. Análise de agrupamento para documentos. Modelagem de tópicos. Modelagem preditiva apoiada em texto: classificação e regressão. Introdução ao processamento natural da linguagem. Web scraping e web mining. Formato de dados de API Web: XML e JSON. Introdução ao exame e manuseio de XML e HTML. Mecanísmos de acesso baseado em DOM e SAX. Linguagem de consulta Xpath. Estratégias para extração serial de dados HTML. Automação de web scraping.

**OBS (1): ao assinalar a opção CH em EAD, indicar a carga horária que será à distância.*

Documento assinado eletronicamente por **PAULO JUSTINIANO RIBEIRO JUNIOR**, **CHEF DEPTO ESTATISTICA**, em 09/10/2019, às 17:38, conforme art. 1º, III, "b", da Lei 11.419/2006.

A autenticidade do documento pode ser conferida aqui informando o código verificador **2201050** e o código CRC **754A4815**.

Art. 9º da Resolução 30/90 – CEPE

Padrão (PD): conjunto de estudos e atividades desenvolvidos fundamentalmente nos espaços de aprendizagem considerados padrão para as modalidades de ensino presencial e de educação à distância (EAD).

Laboratório (LB): conjunto de estudos e atividades desenvolvidos fundamentalmente em espaços de aprendizagem estabelecidos com infraestrutura especializada, tais como laboratórios, oficinas e estúdios.

Campo (CP): conjunto de estudos e atividades desenvolvidos fundamentalmente mediante atividades de campo.

Estágio (ES): conjunto de estudos e atividades desenvolvidos fundamentalmente em ambientes de trabalho mediante estágios regulados pela Lei nº 11.778, de 25 desetembro de 2008.

Orientada (OR): conjunto de estudos e atividades direcionados à vivência na atuaçãoacadêmica e/ou profissional, em seus mais amplos aspectos, desenvolvidos em espaços educacionais internos e/ou externos à UFPR, com a participação direta de docente responsável.

Práticas Específicas (PE): conjunto de atividades de natureza prática, desenvolvidas em ambientes que apresentem restrições ao quantitativo de alunos por docente e que exijam controle rigoroso envolvendo questões de segurança, dignidade, privacidade e sigilo e/ou atenção do docente individualizada ou a pequenos grupos para desenvolvimento do processo de ensino-aprendizagem, com a participação direta do docente responsável.

Estágio de Formação Pedagógica (EFP): conjunto de estudos e atividades desenvolvidas fundamentalmente no âmbito da educação básica, sob a forma de "práticas de docência" e "práticas pedagógicas de organização do trabalho escolar", envolvendo a orientação direta docente em ações que vão desde a intermediação no acordo de colaboração entre a UFPR e os estabelecimentos de ensino, até o acompanhamento sistemático e processual do planejamento, da execução e da avaliação das atividades desenvolvidas pelos licenciandos, o que requer o contato contínuo e presencial do professor nos diferentes campos de estágio e consequentemente a limitação de alunos por turma.

BIBLIOGRAFIA BÁSICA (mínimo 03 títulos)

1. KUMAR, A.; PAUL, A. Mastering Text Mining with R. Packt Publishing, 2016.
2. KWARTLER, E. Text Mining in pratice with R. John Wiley & Sons, Limited, 2017.
3. MINER, G.; ELDER, J.; HILL, T. Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications. Academic Press, 2012.
4. SILGE, J.; RODINSON, D. Text Mining with R: A tidy approach. 1st ed. O'Reilly Media, 2017.

BIBLIOGRAFIA COMPLEMENTAR (mínimo 05 títulos)

1. BIRD, S.; KLEIN, E.; LOPER, E. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, 2009.
2. DANNEMAN, N.; HEIMANN, R. Social Media Mining with R. Packt Publishing, 2014.
3. FELDMAN, R.; SANGER, J. The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press, 2006.
4. MITCHELL, R. Web Scraping with Python: Collecting Data from the Modern Web. O'Reilly Media, 2015.
5. MUNZERT, S.; RUBBA, C.; MEIßNER, P.; NYHUIS, D. Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining. Wiley, 2015.
6. NOLAN, D.; LANG, D. XML and Web Technologies for Data Sciences with R. Springer New York, 2013.