

Atividade 01 - Definição do tema

(Este documento deve ser entregue no Classroom por apenas um membro da equipe)

Equipe

Nome: *DataAnomaly*

Repositório GitLab da equipe: <https://gitlab.com/mab0205/introcd2-novelty-detection>

Membros:

- Martín Ávila, 2274183, login01, martinete.avila@gmail.com, Eng. Computação, UTFPR

Tema

"Novelty Detection: Identificación de Temas Emergentes en Corpora de Noticias"

Objetivos

- *Usar técnicas de processamento de linguagem natural para analisar as notícias.*
- *Quais as diferenças no texto dependendo do tipo de notícia ou fonte?*

Fontes de dados

TAP-DLND 1.0(2018- Corpus): 5.440 documentos (notícias) anotados que manifestam certos critérios para a detecção de novidades. Contém reportagens de notícias da web, ao longo do tempo. Coleta de informações sobre 223 eventos de 10 domínios diferentes.

<https://live.european-language-grid.eu/catalogue/corpus/995/overview/>

Desafios

- *Interpretação dos dados, especificamente a parte de análise exploratória e como vai avaliar os resultados.*
- *Ter uma metodologia de avaliação.*

Submissão:

Além de preencher este documento no Classroom (um componente precisa preencher e submeter), a equipe também deve fazer a submissão pelo GitLab. Neste trabalho, o único documento essencial no repositório GitLab é este documento de definição do tema no formato PDF. Implemente sua solução no diretório "01 - Escolha do tema" que foi criado automaticamente no seu repositório. Na raiz deste diretório, inclua um documento chamado

tema.pdf conforme as especificações acima. Inclua também no diretório todos os documentos e códigos usados pelo grupo no desenvolvimento desta atividade (caso existam). Faça o commit/push para o repositório da equipe (URL informada acima). Certifique-se de que o arquivo README.md do repositório esteja atualizado com os nomes, RAs e logins GitLab dos integrantes do grupo.