**UNIVERSIDADE REGIONAL DE BLUMENAU**

**CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA**

**CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

**Disciplina: Processamento de Linguagem Natural**

**Unidade 4 – Análise Sintática, Modelagem e Extração de Informações**

**Data de entrega: 20/05**

**Equipe:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

A partir dos dados textuais já coletados, cada equipe deverá apresentar e documentar as seguintes tarefas:

1. **Extração de entidades** (PER, ORG, LOC) e/ou **atributos linguísticos** (NOUN, VERB, ADJ etc.), apresentando uma análise estatística básica (entidades mais frequentes, frequência ao longo do tempo etc.), com comentários sobre a importância (ou não) disso para entender a base de dados **(3 pontos).**
2. Aplicação e avaliação qualitativa de abordagens extrativas **– sumarização abstrativa, sumarização extrativa e extração de palavras-chave**. Incluir comentários sobre aplicabilidade e relevância para o conjunto de dados. **(3 pontos)**
3. Proposta e/ou aplicação de **classificação de textos por aprendizado supervisionado**, incluindo avaliação quantitativa e qualitativa das métricas (precisão, acurácia, recall etc.), estratégias para rotulação dos dados e resultados obtidos/esperados. **(4 pontos)**

ideias

1. deduzir o tema principal de um capítulo

2. adivinhar em qual livro está um versículo