**UNIVERSIDADE REGIONAL DE BLUMENAU**

**CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA**

**CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

**Disciplina: Processamento de Linguagem Natural**

**Unidade 1 – Introdução ao Processamento de Linguagem Natural (PLN)**

**Data de entrega: 26/08**

**Equipe:** Renan Fiedler de Oliveira

1. **Tema escolhido (Qual tipo de dado textual? Qual o tema? Descrever brevemente):**

Busca de nomes de ruas e pessoas que tiveram sua história marcada na cidade. O sistema irá ajudar as equipes da Gincana de Blumenau, com objetivo de a equipe ter alguma pista de algum item que deve ser entregue a comissão organizadora, mostrando ao usuário o que a pessoa ou rua tem algum envolvimento. Será usado o site do Adalberto Day, onde contém muito conteúdo sobre a história de Blumenau, o sistema irá receber um nome de alguma pessoa e deve mostrar o texto ou paragrafo onde o nome da pessoa é citado.

**2. Descrever possíveis aplicações dentro de cada tipo de tarefa:**

|  |  |
| --- | --- |
| **Tarefa** | **Descrição** (explicar quais as entradas e saídas esperadas, processos envolvidos, com foco no resultado esperado). |
| Classificação de texto | Classificar os parágrafos ou textos encontrados, mostrando se biografia, história da rua, ou algo que aconteceu com a rua ou pessoa. |
| Extração de informações | Reconhecer entidade nomeada (NER), identificando nome da pessoa e/ou rua, trazendo o texto que contém a informação que pode ser relevante. |
| Recuperação de informações | Motor de busca: Busca de nome ou pistas para a gincana que estão no texto encontrado. |
| Chatbot ou agentes | Agente interativo, pois ele receberá o nome e deve mostrar o texto onde o nome é citado. |
| Outros |  |

**3. Indicar fontes/repositórios de dados relevantes sobre o tema escolhido (sites, textos etc.)**

Blog do Adalberto Day (https://adalbertoday.blogspot.com), para o tema principal.

**4. Quais as possíveis dificuldades e obstáculos da aplicação de PLN ao tema escolhido?**

Imagens entre textos; A própria quebra de linha do site ou palavras, por ter muita customização visual, pode atrapalhar no entendimento do contexto em que o nome da pessoa é citado; Trazer o nome da pessoa onde o texto não tem muita informação ou com texto muito longo.