Resumo 2 – PLN

Na segunda aula de Processamento de Linguagem Natural, foram introduzidos os fundamentos que regem as expressões regulares. Motivado pela procura de padrões textuais para extrair trechos específicos de textos que seguem determinada lógica, o estudo das expressões regulares (regex ou regexp) podem ser definidas como cadeias de texto especiais, em uma linguagem formal, para busca/extração de trechos de texto.

Entre as funcionalidades de expressões regulares introduzidas na aula, encontram-se:

1. **Disjunções [ ]**: expressões responsáveis por incluir faixas de caracteres na consulta

Exemplo: a expressão [a-z] procura na expressão um caractere minúsculo entre a e z;

1. **Negação em Disjunções ^:** o símbolo do acento circunflexo nega uma disjunção

Exemplo: a expressão ^[a-z] procura tudo que não seja um caractere minúsculo entre a e z;

1. **Barra Vertical |**: essa expressão funciona como uma lógica OU entre as expressões.

Exemplo: eu|voce procura na expressão pelas frases “eu” ou “você”;

Existem diversos outros parâmetros e outras expressões regulares que realizam diversas outras procuras em trechos. Durante a aula, foram realizados diversos exercícios práticos através do site <https://regex101.com/>, sendo este responsável por apresentar uma excelente interface para testes em expressões regulares. Nesse site, o professor Jesus Mena aplicou e demonstrou exemplos de todas as expressões aprendidas em aula, elucidando dicas e informando sobre possíveis dificuldades durante as implementações.

Em uma segunda parte da aula, foram introduzidos conceitos sobre os possíveis Tipos de Erros. O erro do Tipo 1 relaciona os exemplos classificados como Falsos Positivos que, em outras palavras, podem ser definidos como exemplos classificados dentro da classe positiva sem que estes pertençam a classe positiva. O erro do Tipo 2 refere-se a exemplos dados como Falsos Negativos, sendo estes definidos por amostras pertencentes a classe positiva mas classificadas erroneamente dentro da classe negativa.

Dentro destes conceitos, foram apresentadas métricas que relacionam os Falsos Positivos, Falsos Negativos, Verdadeiros Positivos e Verdadeiros Negativos, denominados Precision e Recall. Tais métricas definem, respectivamente, a minimização das taxas de falsos positivos e falsos negativos. Existem outras métricas que relacionam Precision e Recall (f1\_score, por exemplo) e podem ser utilizadas para avaliação de modelos de classificação (expressões regulares, de fato, classificam trechos textuais).

Em seguida, foi apresentada na aula conceitos de Máquina de Estados Finita para exemplificar expressões regulares. Um teste com algumas questões foi aplicado em sala de aula com o objetivo de fixar o conhecimento sobre algumas técnicas de expressões regulares. Em um exemplo prático, o professor apresentou um dos primeiros sistemas de chatbot criados: Eliza. O segundo exemplo prático bem conceituado é o de aplicação de expressões regulares para identificação de pessoas importantes dentro de artigos científicos.