Slide 1: Logo Facens

Slide 2: Capa do trabalho

Este trabalho sobre sumarização de textos e a apresentação foi desenvolvida pelos alunos Guilherme Proença Cravo da Costa e Renato Druzian com o auxílio do orientador Prof. Johannes von Lochter.

Slide 3: Agenda

A apresentação será baseada no seguinte cronograma: Introdução, Trabalhos correlatos, Escopo do projeto, Recursos utilizados, Demonstração do Protótipo e, para o encerramento, as Metas para o TCC2.

Slide 4: Introdução

Com a grande volume de informação atualmente, a sumarização de textos ajuda para que textos possam ler lidos mais rapidamente sem a perda do contexto.

Com isso, dados de sites de notícias foram coletados com o método webscrapping, que consiste em estruturar os dados para posterior análise.

Também é necessária a implementação de um algoritmo apto para a sumarização, e coletar datasets prontos, que representam os dados de maneira tabular com linhas e colunas.

Slide 5: Trabalhos correlatos

O estudo feito em Distribuited representations of words and phrases and their compositionality, relata que ao representar palavras similares em um vetor, o aprendizado em linguaguem natural possui melhor desempenho.

Em Neural text summarization: A critical evaluation, é proposto simplificar documentos extensos sem perder o sentido com linguagem natural.

Em A study on the CBOW model's overfitting and stability, a relação semântica entre as palavras através de métodos de representação distribuída.

Slide 6: Escopo do projeto

Com a implementação de um webscrapping para coleta e um algoritmo para a sumarização das notícias, quanto maior possível for o dataset, o resultado será mais preciso.

Por enquanto, aproximadamente, 10 páginas de sites de notícias foram analisadas.

Slide 7: Recursos utilizados

Os recursos utilizados, até o momento, são o Visual Studio Code versão 1.45.1 para editar o código-fonte.

Anaconda versão 1.9.12 para facilitar o gerenciamento e a implementação de pacotes.

E o Python versão 3.7 como linguagem de programação, com o auxílio das bibliotecas BeautifulSoup4, Pandas, Numpy, Tensorflow.

Slide 8: Demonstração do protótipo

(a definir)

Slide 9: Metas para o TCC2

Para encerrar a apresentação, as metas para o TCC2 são: analisar todos os dados obtidos com o TCC1 para evoluir o protótipo.

Ampliar a quantidade de datasets de notícias, visto que, para o TCC1 o número foi reduzido para a análise ser mais efetiva.