1o Trabalho Prático
Processamento de Linguagem Natural

O objetivo deste trabalho é praticar conceitos relacionados a vetores semânticos de palavras, e como utilizar esses vetores em uma aplicação real. Para isso vamos utilizar o projeto Gutemberg (https://www.gutenberg.org/) que disponibiliza acesso livre a uma grande lista de livros. Você deverá baixar os livros listados abaixo, e então realizar uma série de análises, listadas a seguir:

1. Baixe o código do algoritmo skip-Gram, visto em sala, e que está disponível em https://github.com/tmikolov/word2vec. No diretório você encontrará informações de como produzir vetores semânticos para palavras.
2. Construa modelos de linguagem para cada livro, construindo vetores de palavras para cada livro.
3. Construa uma matriz quadrada de distância, onde as dimensões da matriz é a distância entre pares de palavras. A matriz pode ser muito grande, e portanto você deverá propor uma solução amostral.
4. Crie uma forma de comparar as diferentes matrizes de distância. Por exemplo:

$$d_2(\mathbf{A}, \mathbf{B}) = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} (a_{ij} - b_{ij})^2}$$

5. Obtenha a matriz de distância entre os livros, e realize uma análise qualitativa em termos, por exemplo, de gênero e autor.

Você deverá entregar documentação (de 3 a 4 páginas) detalhando os passos realizados, bem como as decisões de projeto. Mencione também os desafios enfrentados durante a implementação do projeto.

A Study in Scarlet by Arthur Conan Doyle	Adventures of Huckleberry Finn by Mark Twain	Anna Karenina by graf Leo Tolstoy	An Inquiry into the Nature and Causes of the Wealth of Nations by Adam Smith	Emma by Jane Austen
The Adventures of Sherlock Holmes by Arthur Conan Doyle	Jonathan Swift____Gulliver's Travels	Crime and Punishment by Fyodor Dostoyevsky	Autobiography of Benjamin Franklin by Benjamin Franklin	Pride and Prejudice by Jane Austen
The Hound of the Baskervilles by Arthur Conan Doyle	The Adventures of Tom Sawyer by Mark Twain	Les Misérables by Victor Hugo	Beyond Good and Evil by Friedrich Wilhelm Nietzsche	Sense and Sensibility by Jane Austen
The Return of Sherlock Holmes by Arthur Conan Doyle	The Jungle Book by Rudyard Kipling	The Brothers Karamazov by Fyodor Dostoyevsky	Il Principe	
The Sign of the Four by Arthur Conan Doyle	Three Men in a Boat by Jerome K. Jerome	The Count of Monte Cristo, Illustrated by Alexandre Dumas	John Locke____Second Treatise of Government	
	Treasure Island by Robert Louis Stevenson	War and Peace by graf Leo Tolstoy	Leviathan by Thomas Hobbes	
	Ulysses by James Joyce		The Life and Adventures of Robinson Crusoe by Daniel Defoe	
			The Republic by Plato	
			Utopia by Saint Thomas More	

A Study in Scarlet by Arthur Conan Doyle

Adventures of Huckleberry Finn by Mark Twain

A Study in Scarlet by Arthur Conan Doyle

Adventures of Huckleberry Finn by Mark Twain

Anna Karenina by graf Leo Tolstoy

An Inquiry into the Nature and Causes of the Wealth of Nations by Adam Smith

Emma by Jane Austen

The Adventures of Sherlock Holmes by Arthur Conan Doyle

Jonathan Swift____Gulliver's Travels

Crime and Punishment by Fyodor Dostoyevsky

Autobiography of Benjamin Franklin by Benjamin Franklin

Pride and Prejudice by Jane Austen

The Hound of the Baskervilles by Arthur Conan Doyle

The Adventures of Tom Sawyer by Mark Twain

Les Misérables by Victor Hugo

Beyond Good and Evil by Friedrich Wilhelm Nietzsche

Sense and Sensibility by Jane Austen

The Return of Sherlock Holmes by Arthur Conan Doyle

The Jungle Book by Rudyard Kipling

The Brothers Karamazov by Fyodor Dostoyevsky

Il Principe

The Sign of the Four by Arthur Conan Doyle

Three Men in a Boat by Jerome K. Jerome

The Count of Monte Cristo, Illustrated by Alexandre Dumas

John Locke____Second Treatise of Government

Treasure Island by Robert Louis Stevenson

War and Peace by graf Leo Tolstoy

Leviathan by Thomas Hobbes

Ulysses by James Joyce

The Life and Adventures of Robinson Crusoe by Daniel Defoe

The Republic by Plato

Utopia by Saint Thomas More

A Doll's House - a play by Henrik Ibsen

A Tale of Two Cities by Charles Dickens

A Christmas Carol in Prose; Being a Ghost Story of Christmas by Charles Dickens

Common Sense by Thomas Paine

Frankenstein; Or, The Modern Prometheus by Mary Wollstonecraft Shelley

Pygmalion by Bernard Shaw	Anne of Green Gables by L. M. Montgomery	Alice's Adventures in Wonderland by Lewis Carroll	Don Quixote by Miguel de Cervantes Saavedra	Moby Dick; Or, The Whale by Herman Melville
The Importance of Being Earnest: A Trivial Comedy for Serious People by Oscar Wilde	David Copperfield by Charles Dickens	Around the World in Eighty Days by Jules Verne	Paradise Lost by John Milton	The Scarlet Letter by Nathaniel Hawthorne
	Great Expectations by Charles Dickens	Candide by Voltaire	Songs of Innocence, and Songs of Experience by William Blake	The Works of Edgar Allan Poe — Volume 1 by Edgar Allan Poe
	Jane Eyre- An Autobiography by Charlotte Brontë	Dracula by Bram Stoker	The Complete Works of William Shakespeare by William Shakespeare	The Works of Edgar Allan Poe — Volume 2 by Edgar Allan Poe
	Oliver Twist by Charles Dickens	Metamorphosis by Franz Kafka	The Divine Comedy by Dante, Illustrated by Dante Alighieri	Walden, and On The Duty Of Civil Disobedience by Henry David Thoreau
	The Picture of Dorian Gray by Oscar Wilde	Peter Pan by J. M	The Iliad by Homer	
	The Secret Adversary by Agatha Christie	The Mysterious Affair at Styles by Agatha Christie	The King James Version of the Bible	
	Wuthering Heights by Emily Brontë	The Strange Case of Dr. Jekyll and Mr Hyde	The Tragedy of Romeo and Juliet by William Shakespeare	
		The Time Machine by H. G. Wells		
		The War of the Worlds by H. G. Wells		

The Wonderful Wizard
of Oz by L. Frank Baum

Through the
Looking-Glass by Lewis
Carroll