

UNCOVERING INSIGHTS FROM CORPUS DATA THROUGH UTILIZATION OF ACCESSIBLE ANALYTICAL TOOLS

A large and structured collection of written or spoken texts used for linguistic analysis and research.

It provides a rich source of language data for studying language variation, developing language models, evaluating language technologies, and more

Corpora: My Friend The Murderer, A Study of The Scarlet, The Mystery of Cloomber Metadata: Title, Genre, Publishing Date

- Project Gutenberg for the text files
- Voyant tool for data visualization

In all 3 of the books, 'Said' was the most used word and 'man' came second: which means most of the charcters were men, quite contradictory to modern times literature, and more.

https://aysa2018.github.io/corpusassignment/

What is Corpus?

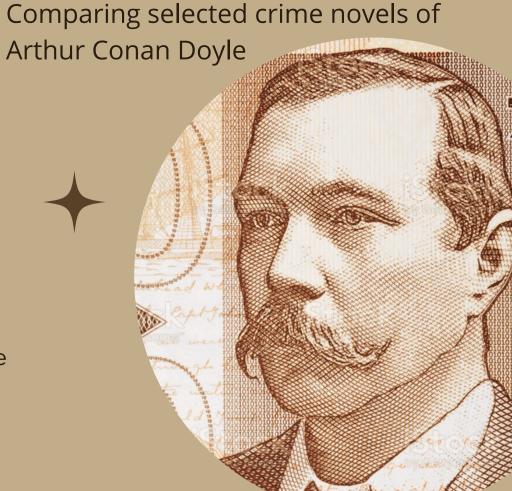
Importance of Corpus

Metadata of selected corpora

Tools used

Interesting trends

Find the whole project at



Name: Aysa Binte Masud

Course: IM-UH 1511 Introduction

to Digital Humanities

Instructor: David Joseph Wrisley, Professor of Digital Humanities

جامعة نيوپورك أبوظي

