

Faculty of Computer Science

Software Proposal Document for project: Forensics Linguistics and Authorship Attribution using Stylometry

Yousef Mohamed - 184367
Supervised by: Dr. Wael Gomaa

Abstract

The main purpose of this project is to study stylometric features and apply those features using machine and deep learning techniques to identify writing authorship in disputed writings, and conduct linguistic analysis on the text and return the findings according to the user's requirements.

1 Introduction

1.1 Background

In written text, each person has a unique style of writing similar to a fingerprint, and in order to identify that writing style linguists have devised methods and researched features that allows those styles to be differentiated and identified from one another.

1.2 Motivation

The process of identifying a writing's features is a difficult task that involves excessive manual labour and a very long time analysing each sentence and clause several times to extract all of the features in the corpus.

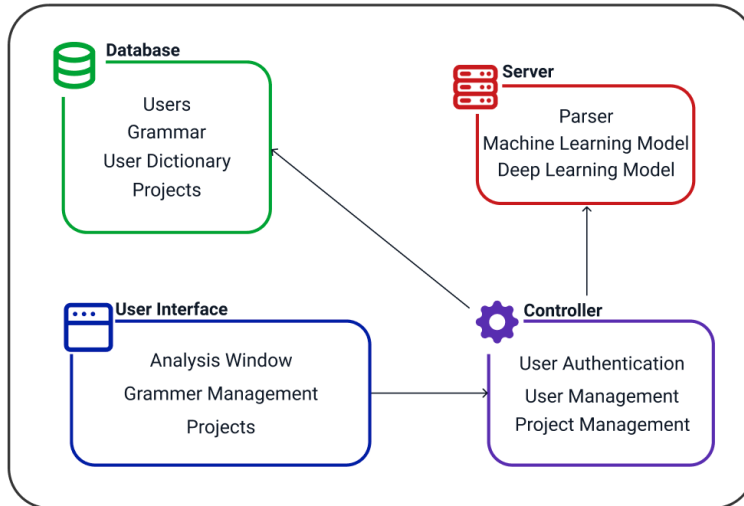
1.3 Problem Definitions

The main problems are:

- Converting the stylometric features to machine readable form.
- Applying those features to machine and deep learning models for analysis.

- Providing a functional interface that allows for complete utilization of the system and its features.

2 Project Description



2.1 Scope

The project aims to be used by linguistic professionals to assist them in writing analysis and feature extraction/validation.

2.2 Project Overview

- Document and project storage.
- Analysis server.
- User interface.

3 Project Management and Deliverable

3.1 Tasks and Time Plan

