Requirements Specification Document

for

author identification project

ariel university

by Eviatar Nachshoni, Yosef Danan, Nir Son

under supervision of Prof. Lee-Ad Gottlieb

Introduction

Purpose

The purpose of this project is to train a machine learning model to identify the author of a given text, out of a close set of authors.

Project Scope

On our project, we try to fit different known models to our data. We focus on two parts of the process:

- Feature selection
- Model picking

The specific implementation of the models it's outside the scope of the project, as we will be using the implementations from existing libraries.

References

Ideas and background were taken from the following papers:

"Computational Methods in Authorship Attribution", Moshe Koppel and Jonathan Schler, Bar-Ilan University. Shlomo Argamon, Illinois Institute of Technology.

"AUTHORSHIP ATTRIBUTION OF RESPONSA USING CLUSTERING", Yaakov HaCohen-Kerner & Orr Margaliot, Jerusalem College of Technology – Lev Academic Center.

"Cross-domain Authorship Attribution: Author Identification using Char Sequences, Word Uni-grams, and POS-tags Features", Yaakov HaCohen-Kerner, Daniel Miller, Yair Yigal, and Elyashiv Shayovitz, Jerusalem College of Technology, Lev Academic Center.

Overall Description

Product Perspective

This project will yield a model and it results analysis. The model can later be integrated into an app with UI, if the result will prove sufficiently good.

Product Features

- A paper summarizing the prosses of feature selection and models training and testing, with the results of the different models.
- The code used.

Operating Environment

The code part of the project is to be run as a standard python 3 project.

Design and Implementation Constraints

Our main limitation is processing power – the entire prosses is to be done on a standard laptop. This may limit out ability to run some complicated models or use many features.

System Features

Results analysis paper

Description

A paper describing the prosses of feature and model selection, including failed attempts. The paper should describe the features and model used, and the result.

Functional Requirements

- For each model and features pair, the paper will present the results – accuracy, precision, recall.
- For each model and features pair, the paper will present the main problems of the results.

Code

Description

All the code used to prepresses the data, generate the different kind of features, train the models, and test them.

Functional Requirements

- The code will provide an easy way to run each of the models
- The code will provide methods to present the results in text / tables / graphs
- The code should include a variety of models, with special attention to neural networks.