

# Software requirement specification document for project Forensics Linguistics and Authorship Attribution using Stylometry

Yousef Mohamed - 184367  
Supervised by: Dr. Wael Gomaa

January 23, 2021

## 1 Introduction

### 1.1 Purpose of this document

The purpose of this document is to define the requirements to be met by the system to be classified as successful and functional.

### 1.2 Scope of this document

The system is wholly engineered and developed by me, with the supervision of Dr. Wael Gomaa. The users are the professors of MSA University's Faculty of Languages.

### 1.3 Overview

The system will take a collection of texts, formatted and organized by author, to be analysed by the system through a set of stylometric and linguistic features, and will return the extracted features embedded in the text with a statistical report of the analysis. The system will also require a selection of predefined grammar sets or new grammar sets defined by the user.

### 1.4 Business Context

The MSA University Faculty of Languages is the organization supporting and contributing to the project.

## **2 General Description**

### **2.1 Product Functions**

1. Create new project.
2. Upload texts.
3. Select from predefined grammar sets.
4. Create new grammar sets.
5. Extract stylistic and linguistic features.
6. Statistical analysis.
7. Highlight the extracted features.
8. Modify analysis results.
9. Login
10. Logout
11. Signup

### **2.2 User Characteristics**

The users will be linguistic professionals that can verify the output of the system, and are able to evaluate and verify the results of the analysis.

### **2.3 User Problem Statement**

The main problem the system is facing is the time consuming labour involved in extracting those linguistic features and their analysis.

### **2.4 User Objectives**

- Identify features of the text.
- Modify the analysis results.
- Identify the disputed text's author.
- Create custom grammar for new features.

### 3 Functional Requirements

Function Name	Create New Project
Description	The user will create a new project that will host the documents of the same context to be analysed.
Criticality	This requirement is the opening point of the system, so the system cannot work without it.
Technical issues	None.
Risks	None.
Dependencies with other requirements	None.
Precondition	Starting state.
Post-Condition	Awaiting documents to be uploaded.

Requirement 1: Create New Project

Function Name	Upload Texts
Description	The user will upload a collection of texts with each group of texts put in a separate folder with the label of the folder being the name of the author, and the disputed text if available.
Criticality	This requirement is the opening point of the project, so the system cannot work without it.
Technical issues	Detecting the user uploaded the texts in the proper format.
Risks	None.
Dependencies with other requirements	'Create New Project'
Precondition	A new project is created.
Post-Condition	Awaiting the features to be explored.

Requirement 2: Upload Texts

Function Name	Select from predefined grammar sets
Description	The user will select a number of grammar sets that the text will be analysed for.
Criticality	System can't advance without it.
Technical issues	None.
Risks	None.
Dependencies with other requirements	None.
Precondition	Texts have been uploaded.
Post-Condition	Awaiting analysis results.

#### Requirement 3: Select from predefined grammar sets

Function Name	Create new grammar sets
Description	The user will be presented with an interface that will allow them to define their own grammar to apply on the corpus. The generated grammar will also be saved to the user's account.
Criticality	Optional.
Technical issues	Ensuring that the grammar is of proper and valid format.
Risks	None.
Dependencies with other requirements	None.
Precondition	Grammar sets have been selected.
Post-Condition	Awaiting analysis of the corpus.

#### Requirement 4: Create new grammar sets

Function Name	Extract stylistic and linguistic features
Description	The system will receive the user's input and proceed to make the required analysis and return the result.
Criticality	Critical, system can't function without it.
Technical issues	None.
Risks	None.
Dependencies with other requirements	'Create New Project', and 'Upload texts', 'Select from predefined grammar sets'
Precondition	Grammar is selected.
Post-Condition	Feature extraction is complete and returned to the user.

#### Requirement 5: Extract stylistic and linguistic features

Function Name	Statistical analysis.
Description	The system will apply statistical analysis to the extracted features received from the parser.
Criticality	Critical, system can't function without it.
Technical issues	None.
Risks	None.
Dependencies with other requirements	'Extract stylistic and linguistic features'
Precondition	Features are extracted.
Post-Condition	Statistical analysis is applied and ready for modification.

#### Requirement 6: Statistical analysis

Function Name	Highlight the extracted features.
Description	The system will highlight the extracted features in the workspace section of the project and offer methods of editing of the results.
Criticality	Not critical.
Technical issues	None.
Risks	None.
Dependencies with other requirements	'Extract stylistic and linguistic features'
Precondition	Features are extracted.
Post-Condition	System awaiting features modification.

#### Requirement 7: Highlight the extracted features

Function Name	Modify analysis results.
Description	The user can hover over a word in the workspace (highlighted or not) and edit its feature association.
Criticality	Not critical.
Technical issues	None.
Risks	None.
Dependencies with other requirements	'Highlight extracted features'
Precondition	Features are highlighted in the workspace.
Post-Condition	Modifications are applied to the workspace and the statistical analysis.

#### Requirement 8: Modify analysis results

Function Name	Login.
Description	User is logged in to system using their credentials.
Criticality	Critical.
Technical issues	None.
Risks	None.
Dependencies with other requirements	None.
Precondition	User has no access to the system.
Post-Condition	User is greeted with the home screen.

#### Requirement 9: Login

Function Name	Logout.
Description	User is logged out of the system and is greeted with the homepage.
Criticality	Critical.
Technical issues	None.
Risks	None.
Dependencies with other requirements	None.
Precondition	User has access to the system.
Post-Condition	User is greeted with the Login screen.

#### Requirement 10: Logout

Function Name	Sign-up.
Description	User creates a new account using their credentials.
Criticality	Critical.
Technical issues	None.
Risks	None.
Dependencies with other requirements	None.
Precondition	User has no account registered in the system.
Post-Condition	User account is registered and they are greeted with the system's homepage.

#### Requirement 11: Sign-up

# 4 Interface Requirements

## 4.1 User Interfaces

### 4.1.1 GUI

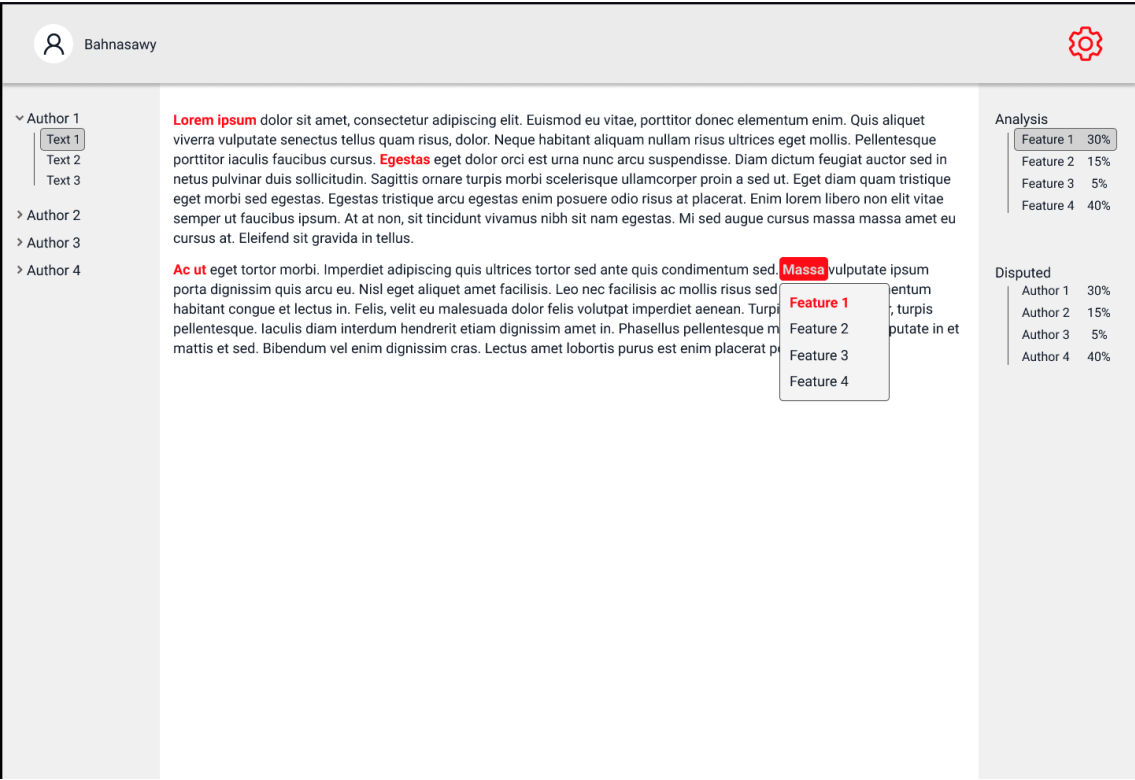


Figure 1: Screen 1

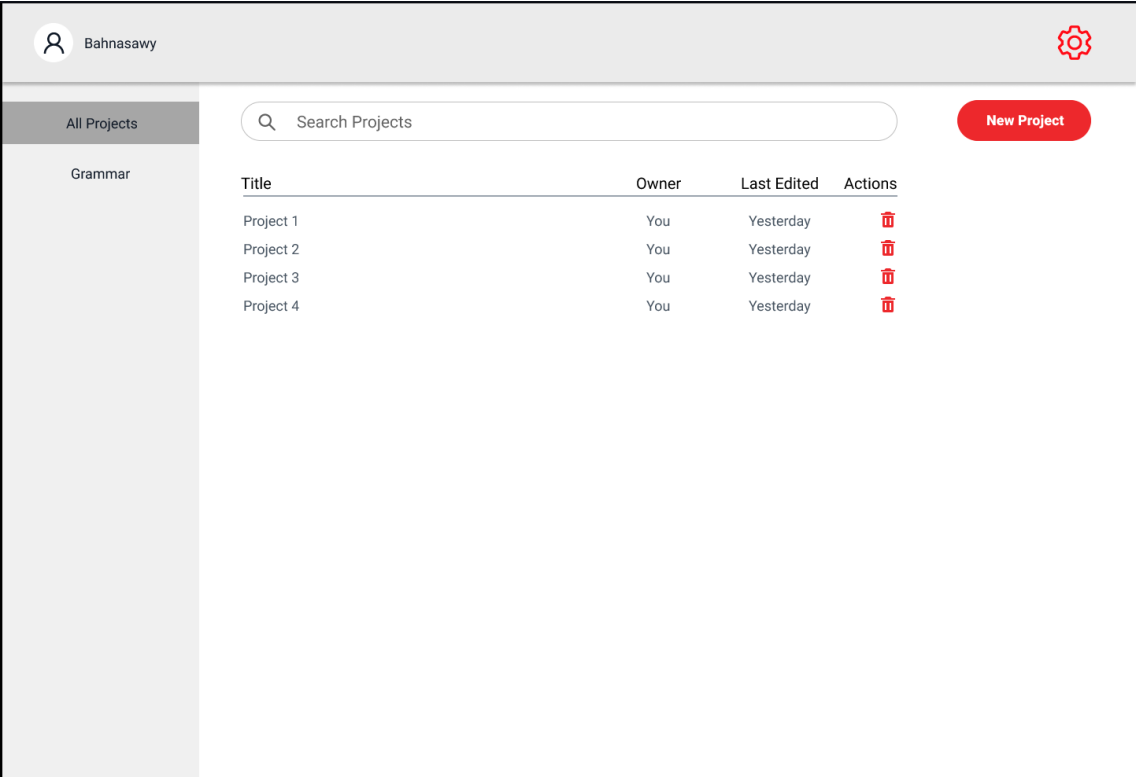


Figure 2: Screen 2

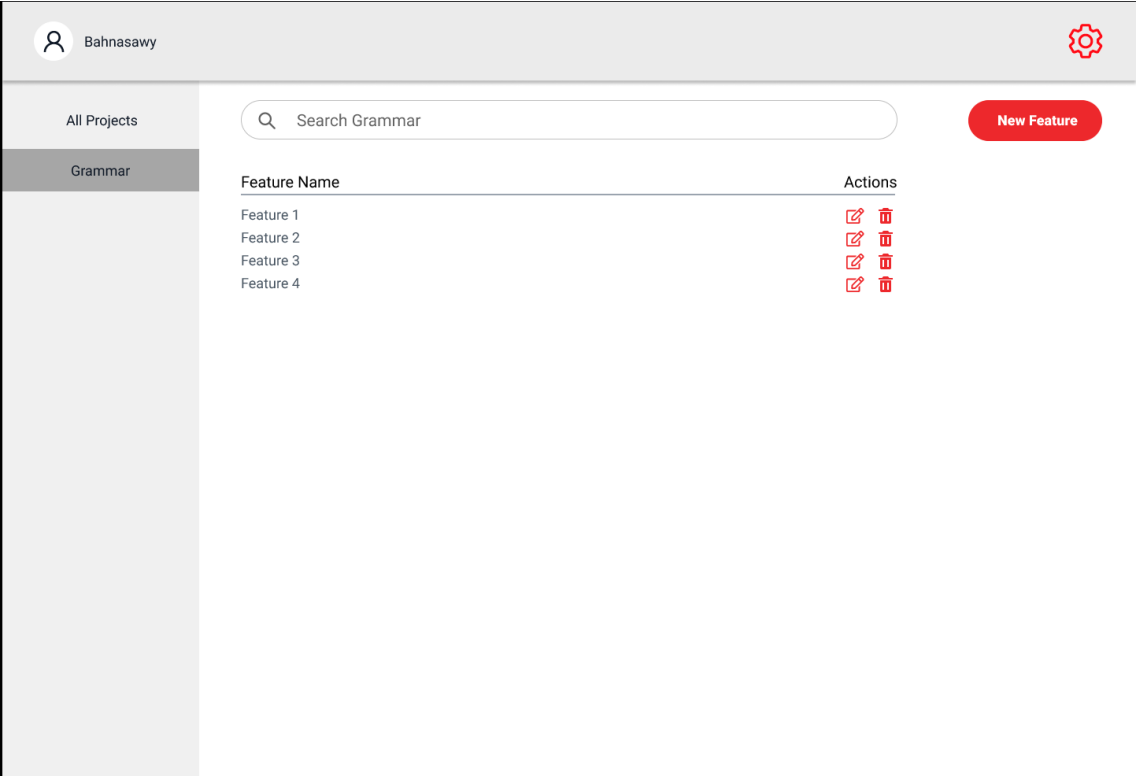


Figure 3: Screen 3



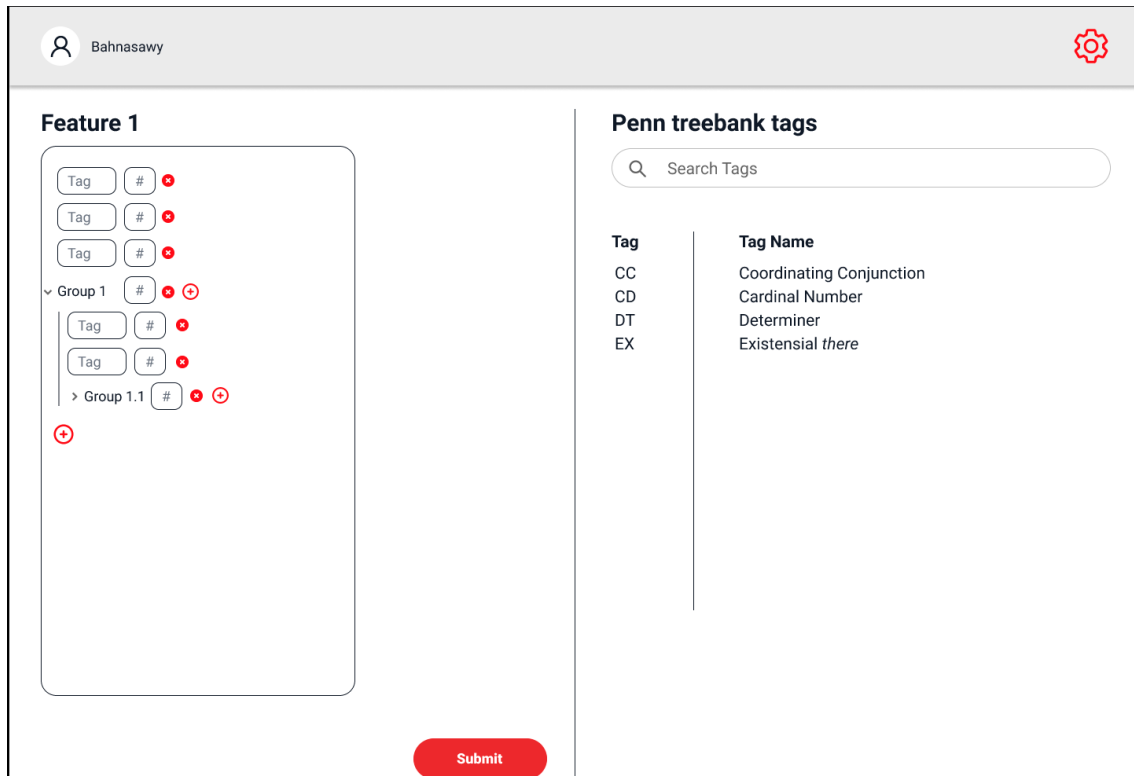


Figure 4: Screen 4

## 5 Design Constraints

No constraints provided.

## 6 Other non-functional attributes

### 6.1 Security

The system will use JWT (JSON Web Token) for user authentication.

### 6.2 Reliability

The system will be deployed as web services on AWS which offers a 99.99% reliability assurance.

### 6.3 Maintainability

The system will be utilizing the micro-services architecture ensuring smaller and more maintainable code.

### 6.4 Portability

The system will be web based, so it is absolutely portable.

## 6.5 Extensibility

The system is using the micro-services architecture, so it is easily extensible.

## 6.6 Re-usability

The system is using the micro-service architecture, so any bit of code can be reused.

## 7 Operational Scenarios

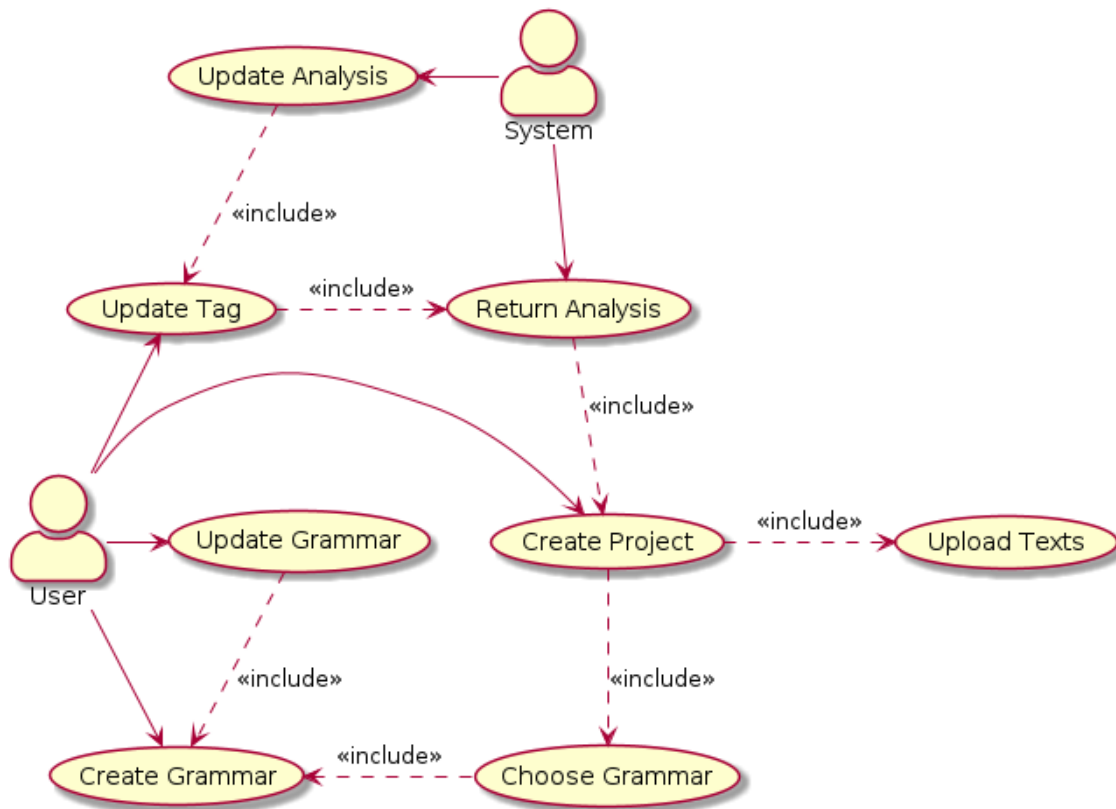


Figure 5: Use case Diagram

## 8 Preliminary Schedule Adjusted

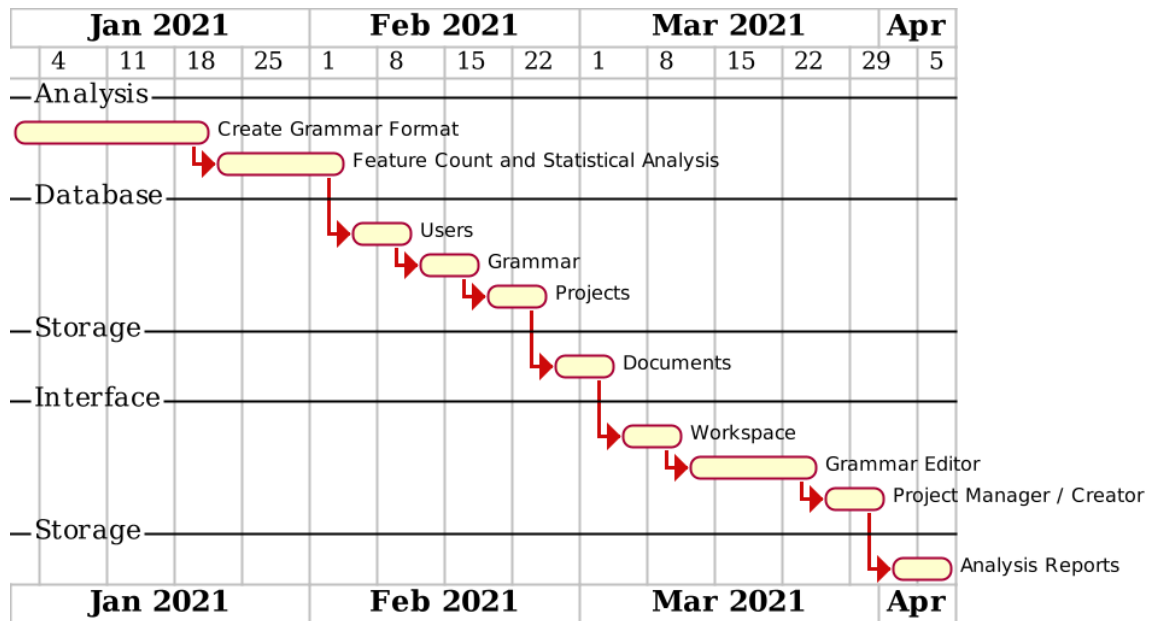


Figure 6: Gantt Chart