Software Requirements Specification

for

<Keyword Clustering Tool for SEO Optimization>

Version 1.0 approved

Prepared by <Andrii Martynenko>

<Kharkiv National University of Radioelectronics>

<June 2025>

Table of Contents

1. Introduction 1

1.1 Purpose 1

1.2 Document Conventions 1

1.3 Intended Audience and Reading Suggestions 1

1.4 Product Scope 1

1.5 References 1

2. Overall Description 2

2.1 Product Perspective 2

2.2 Product Functions 2

2.3 User Classes and Characteristics 2

2.4 Operating Environment 2

2.5 Design and Implementation Constraints 2

2.6 User Documentation 2

2.7 Assumptions and Dependencies 3

3. External Interface Requirements 3

3.1 User Interfaces 3

3.2 Hardware Interfaces 3

3.3 Software Interfaces 3

3.4 Communications Interfaces 3

4. System Features 3

4.1 Load CSV File 3

4.2 Query Clustreing 4

4.3 Output of results 4

4.4 Interaction with results 4

5. Other Nonfunctional Requirements 4

5.1 Performance Requirements 4

5.2 Safety Requirements 4

5.3 Security Requirements 4

5.4 Software Quality Attributes 4

5.5 Business Rules 5

6. Other Requirements 5

Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Date** | **Reason For Changes** | **Version** |
| Initial release | June 2025 | Initial version completed | 1.0 |
|  |  |  |  |

# Introduction

## Purpose

This document defines the functional and non-functional requirements for a desktop keyword clustering application designed to automate SEO analysis. The program processes large sets of search queries, groups them by meaning, and visualizes clustering results.

## Document Conventions

Textual requirements are marked as "FR" for functional and "NFR" for non-functional. Priorities: High (H), Medium (M), Low (L).

## Intended Audience and Reading Suggestions

This document is intended for developers, researchers, students, instructors, and SEO specialists. It is recommended to start with Sections 2 and 4.

## Product Scope

The application is an SEO-support tool that significantly reduces the time required to process a semantic core using unsupervised machine learning methods (KMeans).

## References

* scikit-learn documentation
* TF-IDF & PCA theory
* Google Search Console recommendations

# Overall Description

## Product Perspective

This is a standalone desktop application, independent of external servers. It may be part of a broader SEO automation system.

## Product Functions

* Load CSV file
* TF-IDF text vectorization
* Query clustering (KMeans)
* Cluster name generation
* Chart creation: PCA and frequency histogram
* Search and copy interface

## User Classes and Characteristics

SEO specialists - basic UI operation skills

## Operating Environment

* OS: Windows 10+
* Python 3.10+, .exe version for end users

## Design and Implementation Constraints

* CSV format must include "Keyword" and "Search Volume" fields
* Language: English

## User Documentation

* Built-in interface instruction
* PDF manual (optional)

## Assumptions and Dependencies

* User provides a properly structured file
* Data is in English

# External Interface Requirements

## User Interfaces

* Input fields for file path and number of clusters
* Buttons: "Browse", "Cluster", "Copy"
* Text field with clustering results
* Charts displayed in the right pane of the interface

## Hardware Interfaces

No specific hardware requirements

## Software Interfaces

Libraries used: scikit-learn, matplotlib, pandas, CustomTkinter

## Communications Interfaces

None (offline application)

# System Features

## Load CSV File

REQ-1: The system shall allow the user to load a CSV file containing keywords.

## Query Clustreing

* REQ-2: The system shall perform TF-IDF text vectorization.
* REQ-3: The system shall perform clustering using the KMeans algorithm.

## Output of results

* REQ-4: The program shall display clustering results in both text and graphical form.
* REQ-5: Cluster names shall be generated automatically.

## Interaction with results

* REQ-6: The user shall be able to search by keyword.
* REQ-7: The user shall be able to copy cluster results.

# Other Nonfunctional Requirements

## Performance Requirements

NFR-1: Must process up to 10,000 keywords in ≤ 30 seconds

## Safety Requirements

NFR-2: Validate that input data is present before processing

## Security Requirements

NFR-3: Not applicable, as the application does not process personal data

## Software Quality Attributes

* Usability: clean and intuitive interface
* Reliability: file format validation
* Portability: .exe launch capability

## Business Rules

All operations are within a single session; no automatic data saving

# Other Requirements

* Extensibility: allow for additional clustering algorithms
* Future support for multilingual datasets

Appendix A: Glossary

* TF-IDF - Term Frequency-Inverse Document Frequency vectorization
* PCA - Principal Component Analysis
* KMeans - unsupervised clustering algorithm

Appendix B: Analysis Models

* UML Class Diagram
* Use-case Diagram

Appendix C: To Be Determined List

* TBD-1: Add support for XLSX format
* TBD-2: Enable report saving