**Requirements catalog**

System for analyzing words connections statistics in given text

System for preparing punctuation marks statistics for given text

Project for the course Development of Information Systems

Alonso Everest, Patrik Homola, Kamil Pecela

Table of Contents

[1. Introduction 3](#__RefHeading___Toc1533_2572116372)

[1.1. The purpose of this requirements catalog 3](#__RefHeading___Toc1535_2572116372)

[1.2. Scope of system use 3](#__RefHeading___Toc1537_2572116372)

[1.3. Glossary 3](#__RefHeading___Toc1539_2572116372)

[1.4. Overview of the following chapters 3](#__RefHeading___Toc1541_2572116372)

[2. General description 5](#__RefHeading___Toc1543_2572116372)

[2.1. System perspective 5](#__RefHeading___Toc1545_2572116372)

[2.2. Functions of the system 5](#__RefHeading___Toc1547_2572116372)

[2.3. Users characteristics 5](#__RefHeading___Toc1549_2572116372)

[2.4. General restrictions 5](#__RefHeading___Toc1551_2572116372)

[2.5. Prerequisites and dependencies 5](#__RefHeading___Toc1553_2572116372)

[3. Specific requirements 6](#__RefHeading___Toc1241_3193223655)

[3.1. Opening the system 6](#__RefHeading___Toc1243_3193223655)

[3.2. Analysis session in both systems 6](#__RefHeading___Toc1245_3193223655)

[3.3. Query session in first system 6](#__RefHeading___Toc1247_3193223655)

# [Introduction](#__RefHeading___Toc1006_2572116372)

## [The purpose of this requirements catalog](#__RefHeading___Toc1008_2572116372)

This document presents a summary of all requirements for two information systems. First one produces words connection statistics for given text and makes graphs based on users queries. Second one produces statistics of punctuation marks for given text and makes graphs illustrating the outcomes. It was created as a project within the subject “Development of Information Systems” at the Faculty of Mathematics, Physics and Informatics of the Comenius University in Bratislava. The document is created based on the requirements that were written down at the meeting with the client and in the subsequent communication. It is intended for all persons involved in the development of the system, its administration and use. It also serves as a binding agreement on the functionality of the system between the client and the creators.

## [Scope of system use](#__RefHeading___Toc1049_2572116372)

The goal is to develop two applications. First one will be used as a tool in a search for signs of aphasia in texts based on statistics of the closest connections between words. The system will analyze given input file and produce statistics of every word connection in it. Statistics will be stored in database. On the basis of this statistics system will produce graphs. Second system will prepare statistics of punctuations marks in texts. The system will analyze given input file for number of punctuation mark and the lengths (measured in the number of words) between any two punctuation marks: longest, shortest and average for given input. On the basis of this statistics system can produce graphs. Application will run independently on personal computers.

## [Glossary](#__RefHeading___Toc1012_2572116372)

* **Text** – text in traditional meaning, saved in one of the computer formats, that user wants to analyze.
* **Input file** – text preprocessed by application: \*.txt file containing text without editorial information transformed into one paragraph without punctuation marks.
* **Analysis session: words connections** – process of transforming text to input file, analyzing (counting) words connections, producing statistics, saving them in non-editable data files. After completing analysis text is removed.
* **Statistics of word connections** – output of analysis session stored on users computer in form of non-editable data files.
* **Analysis session: punctuation marks** – process of transforming text to input file, analyzing (counting) distance between punctuation marks and producing statistics. After completing analysis text is removed.
* **Query session** – process of analyzing statistics of word connections by user with possibility to produce graphs.

## [Overview of the following chapters](#__RefHeading___Toc1076_2572116372)

The following chapters will describe the functionality of the system, general limitations, functional and quality requirements and interface requirements.

# [General description](#__RefHeading___Toc1016_2572116372)

## [System perspective](#__RefHeading___Toc1018_2572116372)

First system will serve for academic scholars searching for possible symptoms of aphasia in texts to analyze any given text, prepare statistics of words connection and make graphs on this basis. Aim of

the system is to store statistics of analyzed texts and enable easy way of querying these statistics to produce graphs. Statistics of connections between words may help to find or define symptoms of aphasia in texts. Second system is aimed at analyzing any given text, preparing statistics of punctuation marks and making graphs on this basis. Statistics of punctuation marks may help to study language from different points of view.

## [Functions of the system](#__RefHeading___Toc1020_2572116372)

First system consists of two parts. First part transforms text given by the user into input file, analyzes it and produces statistics for words connections, then saves the outputs into non-editable data files. After closing analysis session texts and input files are deleted from the memory. Second part serves as an interface to stored data. User queries the database to see parts of statistics for given words and makes graphs on this basis.

Second system will transform text given by the user into input file, analyze it, produce statistics for punctuation marks and illustrate it with the graphs. After finishing analysis texts and input files will be deleted from the memory.

## [Users characteristics](#__RefHeading___Toc1207_2572116372)

User of both systems is an academic researcher presumably without experience in IT. There is only one role and profile of user. User have rights to use all the functionalities of the systems.

## [General restrictions](#__RefHeading___Toc1024_2572116372)

Both systems analyzes only written texts. Output statistics are saved in non-editable files. Applications will not store texts – they will be removed after the end of analysis session. Other restrictions are of copyright nature and are responsibility of the user, who should respect the rights of the analyzed texts authors.

## [Prerequisites and dependencies](#__RefHeading___Toc1209_2572116372)

Application will run in Windows OS. Besides installation it will need free disk space for statistics.

# Specific requirements

## Opening the system

1. User can open both systems by clicking the executable apps.
2. User operates on graphic interfaces.
3. User of first system chooses wether to upload new text and analyze it or work on previous statistics.

## Analysis session

### First system

1. User inputs text from his computer.
2. User describes the text with a title and author and may add optional description: genre and information whether text is transcription of a spoken language and whether author suffers from aphasia.
3. System transforms text into input file into array of words.
4. First system performs analysis by counting number of every word connections in input file.
5. Outcome of analysis are pairs of words (order of them is not important) and number of connections between them. This analysis is performed for every word in the input file.
6. Outcomes in form of database are stored in system directory under the title of the analyzed text.
7. After analysis ends user is informed that outcomes were stored. System goes back to the initial interface.

### Second system

1. User inputs text from his computer.
2. User describes the text with a title and author and may add optional description: genre and information whether text is transcription of a spoken language and whether author suffers from aphasia.
3. System transforms text into input file into array of words.
4. System performs analysis by counting distance between any punctuation marks and the frequency of that distance in the text.
5. System outputs shortest, longest and average distance between any two punctuation marks.
6. Outcomes of the analysis are presented with the graphs.

## Query session in first system

1. User chooses the statistics he wants to investigate by the title of text.
2. User queries chosen statistics in the graphic interface.
3. Key to the query can be word, pair of words, number of connections or any combination of those factors.
4. User can generate graph for his query.
5. After closing analysis session system goes back to the opening interface.