**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 systems 5](#__RefHeading___Toc1547_2572116372)

[2.3. Users characteristics 6](#__RefHeading___Toc1549_2572116372)

[2.4. General restrictions 6](#__RefHeading___Toc1551_2572116372)

[2.5. Prerequisites and dependencies 6](#__RefHeading___Toc1553_2572116372)

[3. Specific requirements 7](#__RefHeading___Toc1241_3193223655)

[3.1. First system: “Aphasia detection” 7](#__RefHeading___Toc1243_3193223655)

[3.2. Second system: “Punctuation mark analysis” 8](#__RefHeading___Toc1245_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, named “Aphasia detection”, produces words connection statistics for given text and makes graphs based on user’s queries. Second one, named “Punctuation mark analysis”, 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 text and produce statistics of every word connection in it. Statistics will be stored in application folder on user’s computer. Based on this statistics system will produce graphs. Second system will prepare statistics of punctuations marks in texts. The system will analyze given text for the lengths (measured in the number of words) between any two punctuation marks: longest, shortest and average. Based on this statistics system can produce graphs. Application will run independently on personal computers.

## [Glossary](#__RefHeading___Toc1012_2572116372)

* **Text** – text in traditional meaning, saved in \*.txt format.
* **Array of words** – text preprocessed by application.
* **Statistics file** – output of analysis stored in application directory on user’s computer in form of data file – format: \*.txt. There is one file for each text. Statistic file contains description of text if the user introduced it while uploading text.
* **Record in log file** – after deleting statistic file proper record is added to log file, with information about deleted file and time of this operation.

## [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, “Aphasia detection”, 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, “Punctuation mark analysis”, 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)s

First system “Aphasia detection” starts with initial graphical user interface displaying three options. First option (A) is to upload text for analysis, second one (B) is to delete file, third one (C) is to view statistics and produce graphs [bar charts].

Ad. A: After clicking button “Upload text”, user is redirected to another GUI. He is uploading the text in form of text document (\*.txt) in utf8. There is additional field for a user to input optional description of uploaded text with disclaimer suggesting options of description (genre and information whether text is transcription of a spoken language and whether author suffers from aphasia). Disclaimer will be visible all the time: “All text within the given file will be analyzed. If you don’t want editorial notes to be analyzed as well, you need to delete it.”. Then user clicks “Go” button and analysis starts. Analysis will firstly preprocess text: it will convert text into simple array of words without punctuation marks. Analysis will count how often any two words appear as a pair regardless of order. This information will be stored in a text file called statistics file and saved in a application directory. User is returned to the initial screen with three choices. He can add another text, run query (see B), view statistics (see C) or end application.

Ad. B: After clicking button “Delete file” user is redirected to interface with list of saved statistics file. He chooses file to delete and execute his decision by clicking proper button. After deleting file new record is added to log file. User can click button “Go to start” (which will redirect him to initial interface) or delete another file.

Ad. C: After clicking button “View statistics and produce graph”, user is redirected to new GUI. There will be a list of statistics files displayed, from which user can choose one to query. Three buttons will allow him to choose searching by word (he must input it in proper bar), by a pair of words (he must input both of them) or search through entire text. In another input bar he can choose the number of outputs. It will be inactive if user chooses to search for a specific pair of words. There will be a check box to produce graphs. User also can input logarithmic scale of displaying graph. At the bottom there will be a button: “See results”. By clicking “See results” user is redirected to interface that displays results: statistics and graphs. By clicking button “Save results”, user can save the results of query and the graph in form of image file in chosen directory. There is a button “Go to start”, which redirects user to initial page.

Second system “Punctuation mark analysis” starts with initial graphical user interface displaying button “Upload text” and button “Start”. Disclaimer will be visible all the time: “All text within the given file will be analyzed. If you don’t want editorial notes to be analyzed as well, you need to delete it.”. User will input author and title of the text. User can also analyze text for chosen punctuation mark or for any punctuation mark. First option will be executed by clicking one of the buttons with punctuation marks. When no such button is clicked, punctuation marks in general will be analyzed. After clicking “Start” analysis will begin. Analysis will firstly preprocess text: it will convert text into simple array of words. Analysis will count the number of words between occurrences of chosen punctuation mark or between occurrences of any punctuation marks. User can also choose the logarithmic scale for display. After the end of analysis results (in the form of a bar graph) will be shown on new page. User can save results in form of image file to chosen directory by clicking button “Save results”. At the bottom of result there will be button “Analyze new text”, which will redirect user to the initial page.

## [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. Users have rights to use all the functionalities of the systems.

## [General restrictions](#__RefHeading___Toc1024_2572116372)

Both systems analyze only written texts. 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

## First system: “Aphasia detection”

1. User opens the system by clicking the executable app.
2. User interacts with the system using a graphical user interface.
3. Initial page of application contains three buttons that represents three main functionalities of the system: 1) uploading and analyzing text; 2) deleting statistics file; 3) displaying statistics and graphs.
4. User can choose to upload and analyze text by clicking button “Upload text”. After doing that he is redirected to new page of application for uploading.
5. During the process of uploading the text disclaimer will be shown: “All text within the given file will be analyzed. If you don’t want editorial notes to be analyzed as well, you need to delete it.”
6. Process of uploading the text by the user consists of choosing text file to upload, adding name of the author and the title and optionally adding description of the text. Text should be in \*.txt file with utf-8 encoding. Description may contain whatever information user considers useful. Process of uploading ends when user clicks the button “Upload”.
7. During the process of uploading the text disclaimer will be shown: “You may describe the text with additional information about genre of the text, whether it is transcription of a spoken language and whether author suffers from aphasia as well as any other useful information”.
8. System starts analysis of the text when the user ends the upload process. At first, the system transforms the text into an array of words, then the system counts how often any two words appear as a pair of consecutive words regardless of their order.
9. System saves results of the analysis in a special directory in the form of \*.txt file with utf-8 encoding. This file will be named after the author and the title of the text. If the user added description in the process of uploading text, this description will be added as a header to output statistics file. Statistics file will contain a list of a pairs of words with number representing the frequency of occurrence for each pair.
10. System will redirect the user to the initial page after the analysis is finished and the statistics file is saved.
11. User can choose to delete the statistics file from the application directory by clicking a “Delete file” button on the initial page. He or she will be then redirected to a new page of application for file removal.
12. Process of deleting the file by the user consists of choosing the file from the list, pressing a button to delete and confirming this decision. After deleting the file, the user will be automatically redirected to the initial page.
13. After deleting the statistics file by the user, system will automatically add a proper record to log file in the system directory. This record will contain information about deleted file and time of operation.
14. User can choose to view statistics of previously analyzed text by clicking “View statistics and produce graph” on the initial page. He or she will then be redirected to a new page of application for viewing statistics.
15. Process of viewing statistics and producing graph consists of several steps. User must choose a statistics file from the list. Then he or she chooses one of three options: searching by a singular word, a pair of words or through entire text. If the user wants to produce a graph, he or she must separately tick a proper checkbox. User can also choose a logarithmic scale for display. The process starts after the user clicks the button “See results”.
16. Results of viewing statistics file and producing bar graph are displayed on a new page of application. At x axis of the graph there will be items representing pairs of words ore one word, depending on the choice made by user. Y axis will show frequency in numbers. User can save outputs of query in form of image format to chosen directory by clicking button “Save results”. At the bottom of this page there is a button “Go to start”, which redirects the user to the initial page.

## Second system: “Punctuation mark analysis”

1. User opens the system by clicking the executable app.
2. User interacts with the system using a graphical user interface.
3. On the initial page of application user is uploading text to be analyzed and describes it with author and title.
4. Uploaded text should be a \*.txt file with utf-8 encoding.
5. Disclaimer will be visible all the time on the initial page: “All text within the given file will be analyzed. If you don’t want editorial notes to be analyzed as well, you need to delete it.”
6. User can choose what kind of punctuation mark should be the object of analysis by clicking one of the buttons with punctuation marks. If none of the punctuation marks is chosen, analysis will be done for occurrences of any punctuation marks.
7. User can input the logarithmic scale of display.
8. Analysis starts after user clicks “Go” button at the bottom of initial page.
9. System starts the process of analysis by transforming text into array of words. Then system will count the number of words between occurrences of chosen punctuation marks or any punctuation marks.
10. After the analysis finishes, results will be displayed on another page. Results will consist of statistics and bar graph. X axis of the graph will represent frequency and y axis will represent distance in words.
11. User can save outputs of query in form of image format to chosen directory by clicking button “Save results”.
12. At the bottom of the result page there will be a button “Analyze new text”, which will redirect user to the initial page.