

# ManuLab System Demonstration

Eugen Antal and Pavol Zajac

Slovak University of Technology in Bratislava, Slovakia

## Overview

ManuLab is a software product for statistical analysis of encrypted historical manuscripts. The document analysis is performed via a chain of *filters* (main building elements). A filter represents any operation realizable on a document transcription divided into a set of pages. The implemented filters allow to change the reading direction, select sub-pages, or a subsection from the document, and calculate several statistics like the index of coincidence, Shannon's entropy,  $n$ -gram frequency, etc.

The software design also includes document visualization, displaying pairs of manuscript pages with corresponding transcriptions.

## Introduction

ManuLab (**Manuscript Laboratory**) is a general framework allowing any researcher to work with different manuscripts in an efficient way, and to apply multiple transformations on the document transcription. The **goal** of this project was to create a framework (application) for document analysis **adapted to historical manuscripts**.

## Requirements

The most important requirement was to enable a **side-by-side manuscript visualization**. This feature allows to display image-transcription pairs. This can be very helpful during a document analysis, especially if it is integrated with a display of analytic results (via filters).

During the analysis of the proposed software we have identified the following design **requirements**:

- 1 Operating system independence.
- 2 Manuscript visualization, including visual data (scanned document), and its transcription.
- 3 Chain of filters. Each filter can do atomic operations on document transcription.
- 4 Adjustable reading direction (both horizontal and vertical).

## Filters - part I

Filter is the **main building element** used to perform any analysis/action on the loaded document. A filter can perform its action **per page** or on the **whole document** transcription (merged pages) depending on the implementation. It is possible to **join** several filters as a **chain of operations**. This chain can be also saved and loaded.

The application is using two types of filters, that

- 1 modify the transcription,
- 2 do not modify the transcription, and are only used in analysis.

## Filters - part II

We have already implemented the following filters:

- $n$ -gram frequency,  $n$ -gram distances,
- index of coincidence,
- Shannon's entropy,
- substitution,
- sub-pages selection,
- changing the read direction,
- pattern search.

The result of the analysis is visualised through pop-up menu for each filter. In most cases, the data can also be exported into a *csv* file for further processing.

## Software design

ManuLab is an open source and multi-platform software, written in *C++*, *Qt* - licensed under Apache License, Version 2.0.

## Main components

The user interface (Figure 1) of the ManuLab software consists of 5 main components:

- Menu (not visible in the figure)
- 1a - selected page (image) of the manuscript,
- 2 - the transcription of the selected page,
- 3 - available filters palette,
- 4 - selected filters palette.

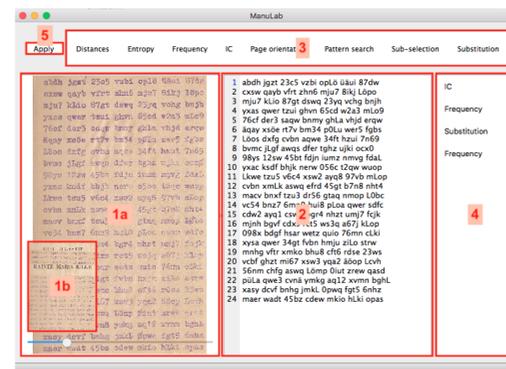


Figure 1: Main components of the UI, displaying a page of the Rilke Cryptogram [1].

## Transcription

The transcription can be also displayed using any custom font installed on the operating system. In Figure 2, the upper part is the original image and the lower part is the transcription using a custom font.

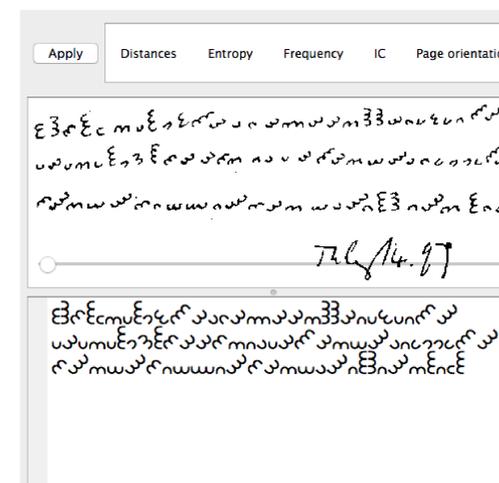
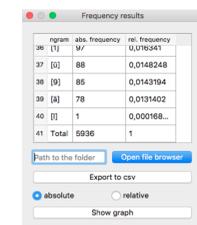


Figure 2: The Dorabella cipher [1].

## Chain of filters

A selected chain of filters with specific settings can be saved to files, thus there is no need to set it up every time. The same manuscript analysis is therefore replicable. Some predefined chains of filters can also be shared between researchers.

Figure 3 shows the frequency analysis result of the Rilke Cryptogram [1].



ngram	abs. frequency	rel. frequency
36 [1]	97	0,016841
37 [0]	88	0,0148248
38 [9]	85	0,0143194
39 [4]	78	0,0131402
40 [1]	1	0,000168...
41 Total	5936	1

Figure 3: Frequency analysis result of the Rilke Cryptogram.

## Source code

The source code is available online at the following GIT repository: <https://bitbucket.org/jugin/manulab.git>.

## References

- [1] Klaus Schmech.  
Klausis Krypto Kolumne.  
<http://scienceblogs.de/klausis-krypto-kolumne>.

## Acknowledgements

This work was partially supported by grant VEGA 1/0159/17.

## Contact Information

- Web: <http://hcportal.eu/manulab.html>
- Email: [eugen.antal@stuba.sk](mailto:eugen.antal@stuba.sk)