Shape

Description automatically generated with medium confidence

School of Management, Economics, Law, Social Sciences,

International Affairs and Computer Science

**The WhatsApp Analyzer**

**Programming with advanced computer languages**

by

Mario Silic

Paul Bruder (16-605-966)

Tobias Klauser (15-608-326)

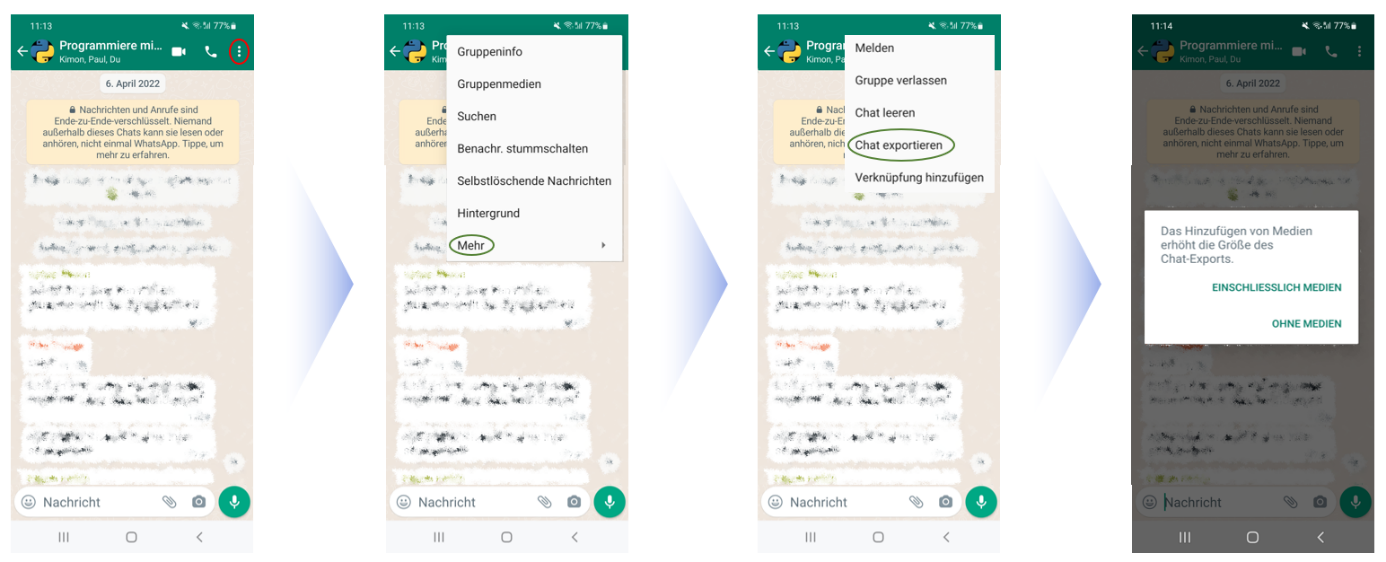
Kimon Ioannidis (16-611-105)

Handed in on

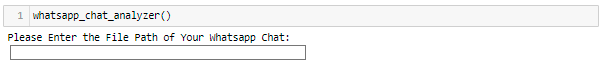
30th June 2022

In school, university, sports clubs, at work or with the family a WhatsApp group exists for every single of these environments. Thousands of messages have been written and build a large unstructured data set. With this project we aimed to create an easy way to analyze the activities within English WhatsApp groups by outputting simple statistics on usage behavior and language. This document intends to give an overview of the project.

As displayed below, WhatsApp offers the option to export a group chat as .txt file.



The chat can be exported with or without files (e.g., pictures, documents etc.). The document should then be saved, and the file path copied. By executing the function *whatsapp\_chat\_analyser()* the path of the WhatsApp chat file can be inserted (see below).



Having inserted the file path, the function requires the user to decide whether only basic statistics should be displayed. (See below).



By entering *Yes,* the program provides first statistics on the:

* Name of the WhatsApp chat file
* Total Messages
* Number of media messages
* Average number of words by message
* Average number of letters by message
* Average number of messages per day

If the user inserts *No* the program will create a new folder and create .png images of extended statistics which include:

* Most active days per month based on messages
* The number of messages over the time of existence of the chat
* Distribution of the messages over 24h hours of a day
* The number of messages by user
* The most active days of the week
* The most frequently used words

Pictures from an exemplary chat can be found in the Appendix. However, statistics for the most active users and the most used words have not been included as picture due to privacy concerns.

Next, the user can enter the number of participants for which he or she would like to generate an analysis of the most common words. A graph per user will then be generated and saved in the same folder.



To see an exact list of the members of the group, the user can type *continue* and a list of all users is displayed. Specific usernames can then be inserted, separated by commas, and an overview is produced for this member of the group chat.

Lastly, the program produces a word cloud for this specific group chat and exports the basic statistics as well as all graphics as a word file. Both are saved in the automatically created folder.

**Libraries used:**

WhatsApp-Analyzer uses several open-source libraries to work properly

* [**Pandas**](https://pandas.pydata.org/)– pandas is an open source, library providing high-performance, easy-to-use data structures and data analysis tools for the Python
* [**Matplotlib**](https://matplotlib.org/)– is a plotting library for the Python programming language and its numerical mathematics extension NumPy.
* [**WordCloud**](https://in.mathworks.com/help/textanalytics/ref/ldamodel.wordcloud.html)– visual representations of words that give greater prominence to words that appear more frequently.
* [**Emoji**](https://pypi.org/project/emoji/)– Emoji for Python.
* **NLTK** – is an open-source compilation of libraries and programs for computational linguistics applications and is distributed under the Apache license
* **whatstk** is a python package providing tools to parse, analyze and visualize WhatsApp chats developed under the society project

Appendix

