6COSC023W – Project Specifications Design and Prototype

Person's Behaviour Analysis with text messages via NLP - BAWT

Student: Mohammad Sabiq Sabry (19219495)

Supervisor: Anastasia Angelopoulou

This report is submitted in partial fulfilment of the requirements for the

BSc (Hons) Computer Science degree

BEng Software Engineering degree

at the University of Westminster.

School of Computer Science & Engineering

University of Westminster

Date: 08/02/2023

Table of contents

List of Figures

List of Tables

1. Aim and Objectives

1.1 Problem Statement

Texting for communication has grown significantly in recent years, particularly with the widespread use of WhatsApp. WhatsApp is the most popular chat application with over 2 billion active users (Ruby, 2022). It is widely used by both businesses and young people.

People check the tone of conversations for a variety of reasons, such as parents wanting to monitor their children's chats or people with conditions like Alexithymia / Autism, which affects their ability to recognise and express emotions (Cherney, 2021).

Because people are not under the pressure of face-to-face social interactions, text messages tend to reveal more honesty in communication. As a result, it is critical to have a system in place to help people with Alexithymia or Autism better understand their communication patterns (Lister-Landman, Domoff, and Dubow, 2017).

1.2 Project Aim

Alexithymic individuals struggle to understand and express their own feelings, as well as to acknowledge and react to the emotions of others so to have a system which will show them how a person behaves will make it easier for one's life. This will also be helpful for the police to have an in-depth understanding of a person's behavioural pattern because in today's world there are so many crimes happening all over the world just via texting.

1.3 Project Objectives

There are 2 objectives I hope to achieve in this paper; one a full-on analysis report of a person's behavioural pattern using Sentiment Analysis. Research shows many people use emojis/gifs to relate to their feelings but sometimes we as people misread it into something else so Sentiment Analysis will portray the positivity level of a person. This will help the people with Alexithymia to understand a person thoroughly without causing distress to one another and it will also help the policia manage the use of improper conduct via text messages.

And two, to propose a system via NLP that, by implementing the base-level implementation, prevents the user from sending unsuitable or improper messages to the participants. Many people in today's world make mistakes when messaging someone important; I hope to present a system where such mistakes can be avoided just by checking the tone of the message.

2. Requirements

2.1 Stakeholders

Individuals with Alexithymia and Autism Spectrum Disorder, as well as law enforcement personnel, will be the primary administrators, users, and beneficiaries of the project.

Individuals suffering from Alexithymia and Autism Spectrum Disorder will benefit from the research project because it will help them understand the emotions behind a text message, enhancing their ability to communicate and form relationships.

The application will be used by law enforcement personnel to judge a person's behaviour, potentially improving the accuracy of their assessments and decision-making processes.

Academic institutions, research organisations and law enforcement agencies may be among the organisations involved in the research's development and implementation. The research project will have an impact on these organisations as well, as it may improve their ability to support individuals with Alexithymia and Autism Spectrum Disorder, as well as make more informed decisions based on behaviour analysis.

Individuals with Alexithymia and Autism Spectrum Disorder, as well as law enforcement personnel and organisations involved in the application's development and implementation are among the project's stakeholders. The application will have an impact on these stakeholders because it will improve their ability to understand and communicate emotions, resulting in better relationships and overall quality of life for people with the mentioned disorders.

2.2 Gathering requirements

In this case, a library called SoAn was used to gather requirements (Social Analysis). Data from WhatsApp messages was extracted using SoAn, which included word frequency, word clouds, TF-IDF, and Sentiment Analysis.

SoAn was used to collect data from the researcher's and a few colleagues' WhatsApp messages who volunteered to share their chats with specific individuals. The messages were then analysed and aggregated to create a thorough examination of the chats (Appendix 2.2).

The data analysis results were summarised using various word frequency and word cloud visualisations, as well as sentiment analysis results. The findings revealed insights into the types of words and emotions commonly expressed in WhatsApp chats, which can be useful in understanding individuals' communication patterns and preferences.

Because the data collected and analysed was limited to the researcher's own chats and a few voluntarily shared chats, the results may not be representative of all WhatsApp users or conversations. To confirm the findings and ensure generalizability, more research with a larger and more diverse sample size is required of which will be collected in due time to the final representation of the project.

2.3 List of project requirements

	Functional Requirements
Essential	● Compatibility with OS ● Python and Libraries like Tensorflow, Pandas, Numpy, etc. are supported ● VSCode ● Google Colab/Jupyter ● Ability to create figures, architectures, documentations - Google Docs/Canva/Draw.io ● Data Backup - Google Drive/Github ● Ability to develop ML/DL/NLP models ● Research Writing Skills
Desirable	● Support for Linux and Windows ● High Performance Hardware - intel i7/M1 or above

	• Massive RAM and Disk Space to manage datasets and development environments
Luxury	• Advanced Data Visualization tools • UI friendly interface

Table 1: Functional Requirements

	Non-Functional Requirements
Essential	• Performance: the application should be able to efficiently process large datasets and perform intensive resource tasks. • Security: the research project should safeguard user privacy and confidential data. • Reliability: the research project should be stable and should not crash when used normally. • Scalability refers to the research project's ability to handle increasing amounts of data and users.
Desirable	• User Experience: the research project should be simple and easy to use. • Flexibility: the research project should be adaptable and integrate with other tools. • Maintenance: the research project should be simple to update and maintain.
Luxury	• Personalization: based on the user's preferences, the research project should make tailored recommendations.

Table 2: Non-Functional Requirements

2.4 Analysis and modelling of requirement

2.4.1 Stakeholder Analysis Diagram

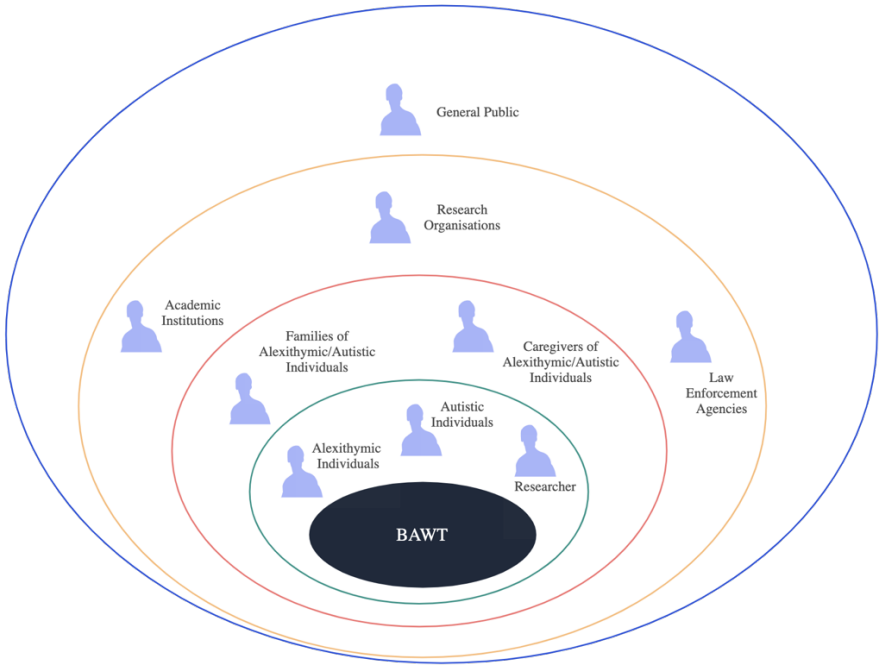


Figure 1: Stakeholder Analysis Onion Model Diagram

As illustrated in the Stakeholder Analysis Model, the project stakeholders include individuals with Alexithymia and Autism Spectrum Disorder, as well as law enforcement personnel and organizations involved in the development and implementation of the application.

2.4.2 Context Diagram

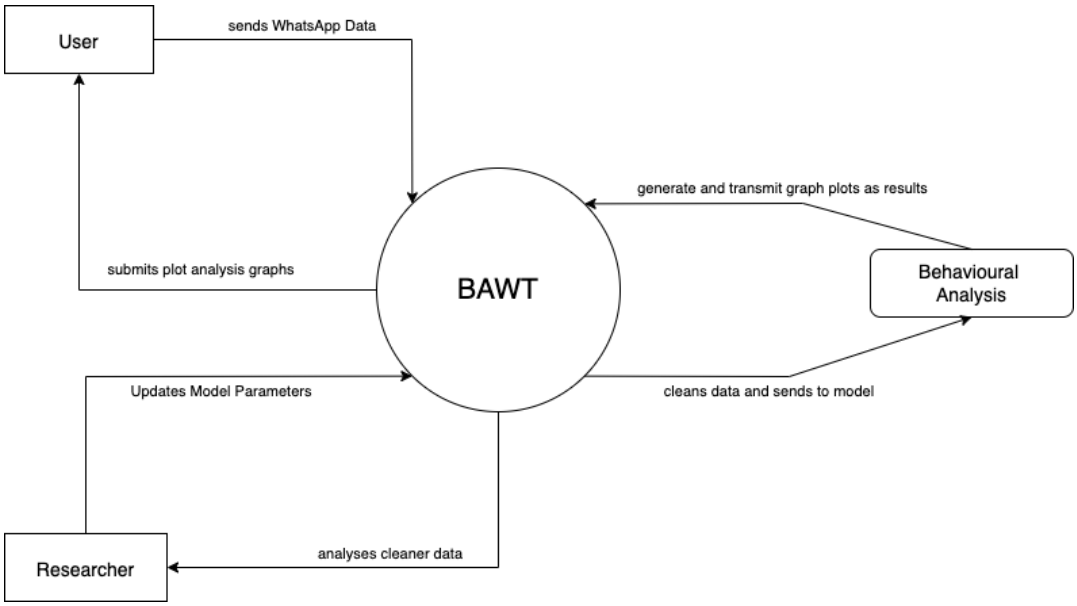


Figure 2: Context Diagram

The diagram illustrated above outlines the boundaries of the system and its interactions. By defining these elements prior to the development phase, the researcher will gain an understanding of the intended flow of information.

2.4.3 Use Case Diagram

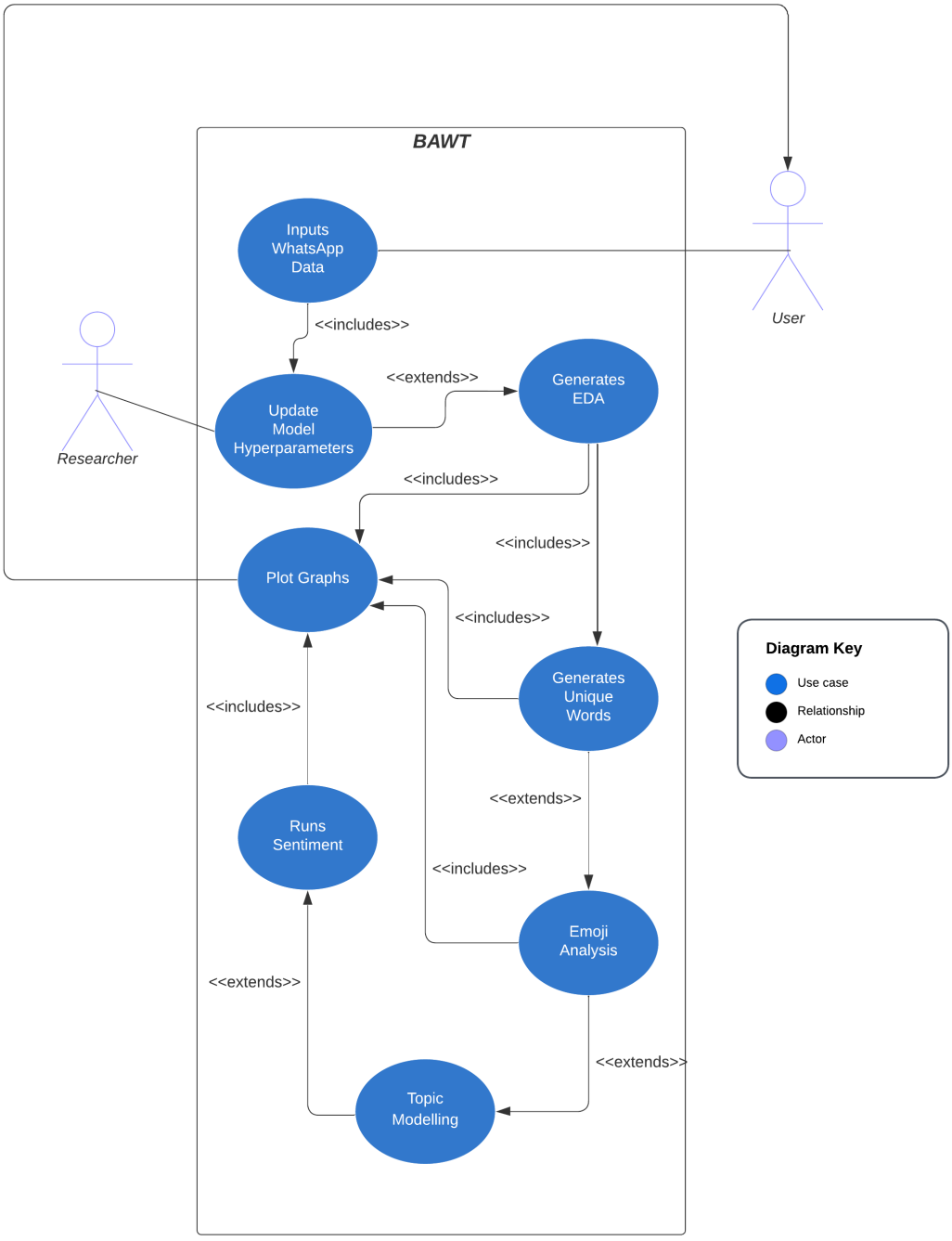


Figure 3: Use Case Diagram

As depicted in the diagram above, the user provides the system with WhatsApp Chat data in the form of .txt files. The Researcher then updates the model hyperparameters to align the variables before running the model.

The data then undergoes a cleaning process and moves on to the Explanatory Data Analysis (EDA) section. This section provides basic analysis charts to provide an overview of the data.

The data then moves on to the Unique Words section, which utilizes the TF-IDF Algorithm to determine the most common and unique words used by the two users and provides charts to illustrate these findings.

The data then proceeds to the Emoji Analyzer, which also utilizes the TF-IDF Algorithm to identify the most commonly used emojis and provides charts to display these results.

Next, using Natural Language Processing (NLP) techniques such as Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF), the system performs Topic Modelling to uncover the most frequently discussed topics between the two users.

Finally, Sentiment Analysis is used to calculate the polarity score of the two users and determine the positivity level (either positive, neutral or negative) in the given period of the given chat.

3. Prototype

Criteria
Prototyping was solely carried out to explore the feasibility of creating the primary research component (Appendix 3).
Discussion of Findings
The researcher conducted basic exploratory data analysis (EDA) and found that the interactions between two users were characterized by their weekly, daily, and hourly chat rate. The researcher also applied the TF-IDF algorithm to identify the most common words used by the two users and visualized these results in charts to improve understanding. Similarly, the researcher used the same algorithm to determine the most commonly used emojis and plotted these results for clarity. Additionally, the researcher utilized topic modeling to identify the most frequently discussed topics by the two users (Appendix 3). During the prototyping phase, the researcher faced some difficulties, including compatibility issues with text exports from iOS devices, which were not well supported by the algorithm designed for Android devices. Another challenge was the limited language support, as the Sentiment Analysis Algorithm currently only works with popular languages, such as English, Dutch, German, and French. However, even though users may speak different languages, they often articulate their messages in English, making it difficult for the algorithm to accurately identify words. Furthermore, the researcher encountered difficulties with an outdated emoji library, making it challenging to recognize new emojis. Lastly, the researcher encountered challenges in assigning different names to users to preserve anonymity, but they hope to find a solution in the future (Appendix 3).

Table 3: Prototype Findings

Here is the video demonstrating the prototype the researcher has made (Appendix 3):

https://youtu.be/m7RXepHpOe4

4. References

sabiqsabry (2023). *sabiqsabry/bawt: People Behaviour via Chat Analysis Model using NLP - BAWT (Behaviour Analysis With Text)*. [online] GitHub. Available at: https://github.com/sabiqsabry/bawt [Accessed 6 Feb. 2023].

Google.com. (2013). *bawt - chat behaviour analysis – Google Drive*. [online] Available at: https://drive.google.com/drive/folders/1aFyn2tiZqEg2P_vIKWao8Pl-s0RggvWP [Accessed 6 Feb. 2023].

Assistive Technology at Easter Seals Crossroads. (2018). *My Emotional Compass App*. [online] Available at: https://www.eastersealstech.com/2018/09/25/my-emotional-compass-app/.

MaartenGr (2020). *GitHub - MaartenGr/soan: Social Analysis based on Whatsapp data*. [online] GitHub. Available at: https://github.com/MaartenGr/soan.

Lister-Landman, K.M., Domoff, S.E. and Dubow, E.F. (2017). The role of compulsive texting in adolescents' academic functioning. *Psychology of Popular Media Culture*, 6(4), pp.311–325. doi:10.1037/ppm0000100.

Silva, A.C.N. da, Branco Vasco, A. and Watson, J.C. (2018). Alexithymia and therapeutic alliance: a multiple case study comparing good and poor outcome cases. *Research in Psychotherapy: Psychopathology, Process and Outcome*, [online] 21(2). doi:10.4081/ripppo.2018.313.

Hobson, H., Hobson, H., Brewer, R. and Bird, G. (2019). *The Role of Language in Alexithymia: Moving Towards a Multiroute Model of Alexithymia*. [online] ResearchGate. Available at: https://www.researchgate.net/publication/330936662_The_Role_of_Language_in_Alexithymia_Moving_Towards_a_Multiroute_Model_of_Alexithymia.

Ruby, D. (2022). *Whatsapp Statistics 2022 — How Many People Use Whatsapp*. [online] demandsage. Available at: https://www.demandsage.com/whatsapp-statistics/.

Bawt Gannt Chart.pdf (2022). *Bawt Gannt Chart.pdf*. [online] Google Docs. Available at: https://drive.google.com/file/d/1uVq9iNuUdM6f-_PQQy1I3l0R4u3gsHXy/view.

5. Bibliography

BMC Blogs. (2020). *Agile vs Waterfall SDLCs: What's The Difference?* [online] Available at: https://www.bmc.com/blogs/agile-vs-waterfall/.

Serani, D. (2014). *The Emotional Blindness of Alexithymia*. [online] Scientific American Blog Network. Available at: https://blogs.scientificamerican.com/mind-guest-blog/the-emotional-blindness-of-alexithymia/.

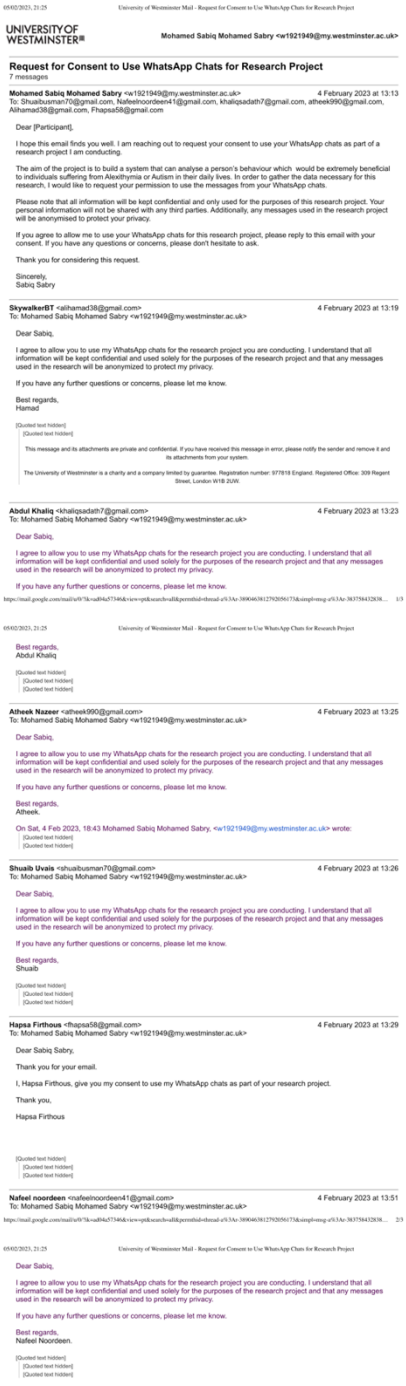
The Berkeley Well-Being Institute. (2019). *Alexithymia: Definition, Symptoms, & Examples*. [online] Available at: https://www.berkeleywellbeing.com/alexithymia.html.

Brent, D. and ryanlight (2021). *The Emotional Mystery of Alexithymia - Beatanxiety*. [online] Beatanxiety. Available at: https://beatanxiety.me/the-emotional-mystery-of-alexithymia/.

Abram, M.D. (2020). *Methods to Integrate Natural Language Processing Into Qualitative Research - Marissa D. Abram, Karen T. Mancini, R. David Parker, 2020*. [online] International Journal of Qualitative Methods. Available at: https://journals.sagepub.com/doi/10.1177/1609406920984608.

6. Appendix

2.2 Gathering Requirements



05/02/2023, 21:25 University of Westminster Mail - Request for Consent to Use WhatsApp Chats for Research Project

UNIVERSITY OF WESTMINSTER

Mohamed Sabiq Mohamed Sabry <w1921949@my.westminster.ac.uk>

Request for Consent to Use WhatsApp Chats for Research Project
7 messages

Mohamed Sabiq Mohamed Sabry <w1921949@my.westminster.ac.uk> 4 February 2023 at 13:13
To: Shuaibusman70@gmail.com, Nafeelnoordeen41@gmail.com, khaliqsadath7@gmail.com, atheek990@gmail.com, Alihamad38@gmail.com, Fhapsa58@gmail.com

Dear [Participant],

I hope this email finds you well. I am reaching out to request your consent to use your WhatsApp chats as part of a research project I am conducting.

The aim of the project is to build a system that can analyse a person's behaviour which would be extremely beneficial to individuals suffering from Alexithymia or Autism in their daily lives. In order to gather the data necessary for this research, I would like to request your permission to use the messages from your WhatsApp chats.

Please note that all information will be kept confidential and only used for the purposes of this research project. Your personal information will not be shared with any third parties. Additionally, any messages used in the research project will be anonymised to protect your privacy.

If you agree to allow me to use your WhatsApp chats for this research project, please reply to this email with your consent. If you have any questions or concerns, please don't hesitate to ask.

Thank you for considering this request.

Sincerely,
Sabiq Sabry

SkywalkerBT <alihamad38@gmail.com> 4 February 2023 at 13:19
To: Mohamed Sabiq Mohamed Sabry <w1921949@my.westminster.ac.uk>

Dear Sabiq,

I agree to allow you to use my WhatsApp chats for the research project you are conducting. I understand that all information will be kept confidential and used solely for the purposes of the research project and that any messages used in the research will be anonymized to protect my privacy.

If you have any further questions or concerns, please let me know.

Best regards,
Hamad

[Quoted text hidden]
[Quoted text hidden]

This message and its attachments are private and confidential. If you have received this message in error, please notify the sender and remove it and its attachments from your system.

The University of Westminster is a charity and a company limited by guarantee. Registration number: 977818 England. Registered Office: 309 Regent Street, London W1B 2UW.

Abdul Khaliq <khaliqsadath7@gmail.com> 4 February 2023 at 13:23
To: Mohamed Sabiq Mohamed Sabry <w1921949@my.westminster.ac.uk>

Dear Sabiq,

I agree to allow you to use my WhatsApp chats for the research project you are conducting. I understand that all information will be kept confidential and used solely for the purposes of the research project and that any messages used in the research will be anonymized to protect my privacy.

If you have any further questions or concerns, please let me know.

https://mail.google.com/mail/u/0/?ik=ad84a57346&view=pt&search=all&permthid=thread-a%3Ar-3890463812792050173&simpl=msg-a%3Ar-3837584432838... 1/3

05/02/2023, 21:25 University of Westminster Mail - Request for Consent to Use WhatsApp Chats for Research Project

Best regards,
Abdul Khaliq

[Quoted text hidden]
[Quoted text hidden]
[Quoted text hidden]

Atheek Nazeer <atheek990@gmail.com> 4 February 2023 at 13:25
To: Mohamed Sabiq Mohamed Sabry <w1921949@my.westminster.ac.uk>

Dear Sabiq,

I agree to allow you to use my WhatsApp chats for the research project you are conducting. I understand that all information will be kept confidential and used solely for the purposes of the research project and that any messages used in the research will be anonymized to protect my privacy.

If you have any further questions or concerns, please let me know.

Best regards,
Atheek.

On Sat, 4 Feb 2023, 18:43 Mohamed Sabiq Mohamed Sabry, <w1921949@my.westminster.ac.uk> wrote:
[Quoted text hidden]
[Quoted text hidden]

Shuaib Uvais <shuaibusman70@gmail.com> 4 February 2023 at 13:26
To: Mohamed Sabiq Mohamed Sabry <w1921949@my.westminster.ac.uk>

Dear Sabiq,

I agree to allow you to use my WhatsApp chats for the research project you are conducting. I understand that all information will be kept confidential and used solely for the purposes of the research project and that any messages used in the research will be anonymized to protect my privacy.

If you have any further questions or concerns, please let me know.

Best regards,
Shuaib

[Quoted text hidden]
[Quoted text hidden]
[Quoted text hidden]

Hapsa Firthous <fhapsa58@gmail.com> 4 February 2023 at 13:29
To: Mohamed Sabiq Mohamed Sabry <w1921949@my.westminster.ac.uk>

Dear Sabiq Sabry,

Thank you for your email.

I, Hapsa Firthous, give you my consent to use my WhatsApp chats as part of your research project.

Thank you,

Hapsa Firthous

[Quoted text hidden]
[Quoted text hidden]
[Quoted text hidden]

Nafeel noordeen <nafeelnoordeen41@gmail.com> 4 February 2023 at 13:51
To: Mohamed Sabiq Mohamed Sabry <w1921949@my.westminster.ac.uk>

https://mail.google.com/mail/u/0/?ik=ad84a57346&view=pt&search=all&permthid=thread-a%3Ar-3890463812792050173&simpl=msg-a%3Ar-3837584432838... 2/3

05/02/2023, 21:25 University of Westminster Mail - Request for Consent to Use WhatsApp Chats for Research Project

Dear Sabiq,

I agree to allow you to use my WhatsApp chats for the research project you are conducting. I understand that all information will be kept confidential and used solely for the purposes of the research project and that any messages used in the research will be anonymized to protect my privacy.

If you have any further questions or concerns, please let me know.

Best regards,
Nafeel Noordeen.

[Quoted text hidden]
[Quoted text hidden]
[Quoted text hidden]

Figure 4: Consent Emails

3. Prototyping

Here is the video once again:

https://youtu.be/m7RXepHpOe4

The example portrayed below was between 2 of the researcher's colleagues:

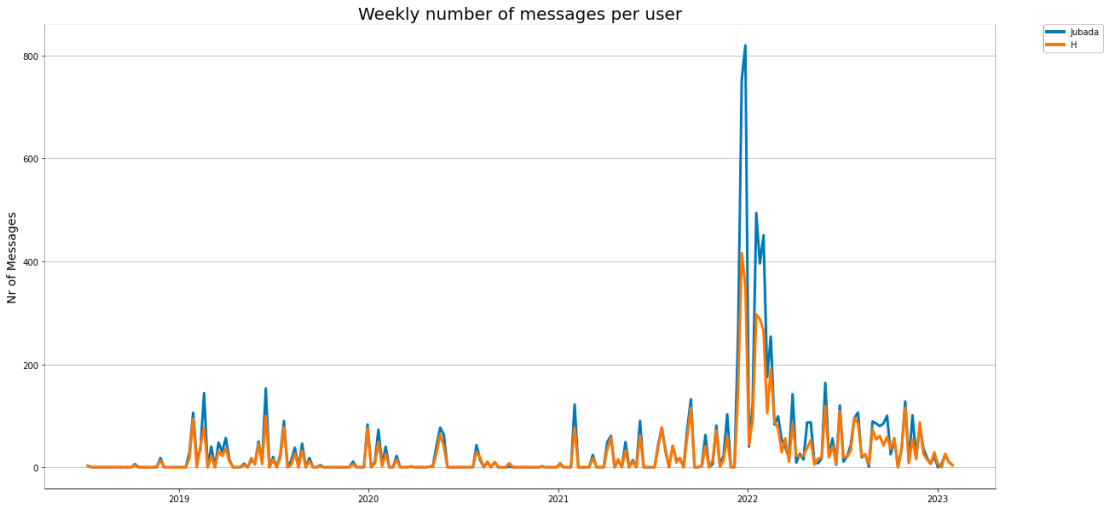


Figure 5: weekly no. of messages

As you can see above with the sample chat used between 2 colleagues of the researcher, it was spiking during the beginning of 2022, which states that something must have been undergone. And as a result, the researcher questioned the user (refer to the image below) in regards to the spike and she mentioned that the other colleague was going through some relationship issues along with her mother getting affected with Covid.

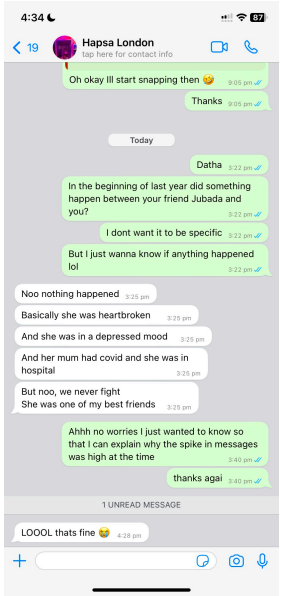


Figure 6: chat as to why spike occurred

Some other findings of the chat were as follows:

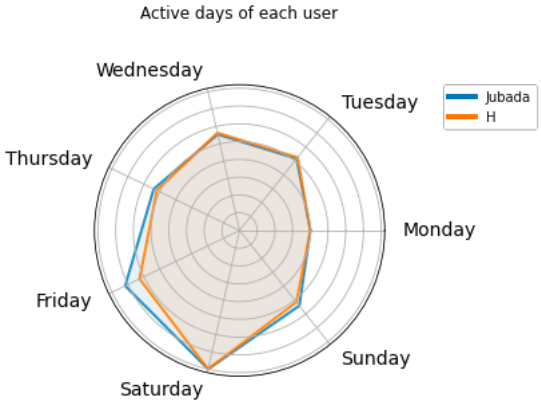


Figure 7: active days of users

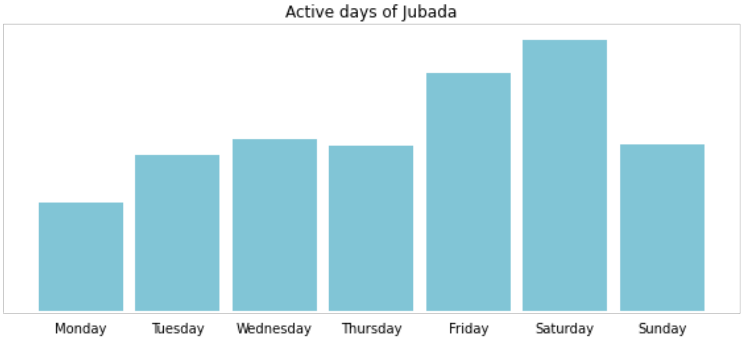


Figure 8: active days of user 1

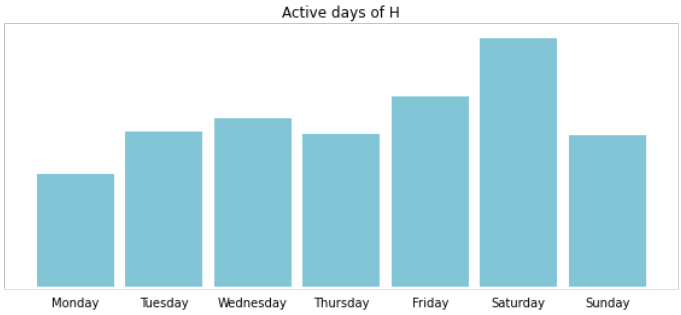


Figure 9: active days of user 2

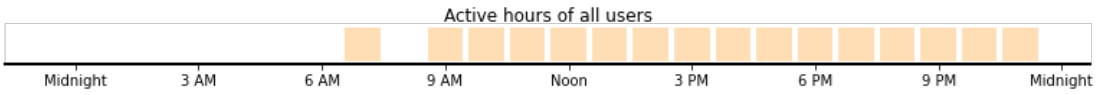


Figure 10: active hours

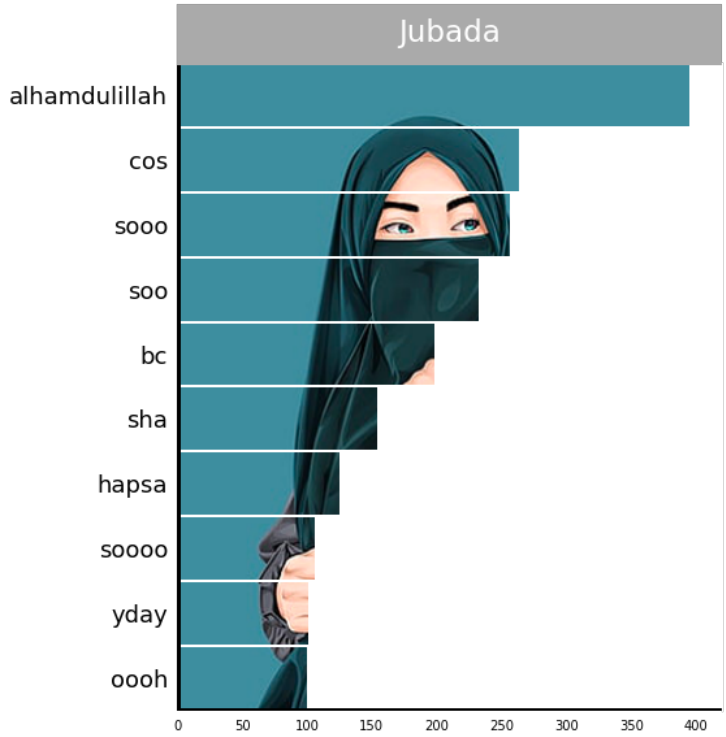


Figure 11: most common words used by user 1

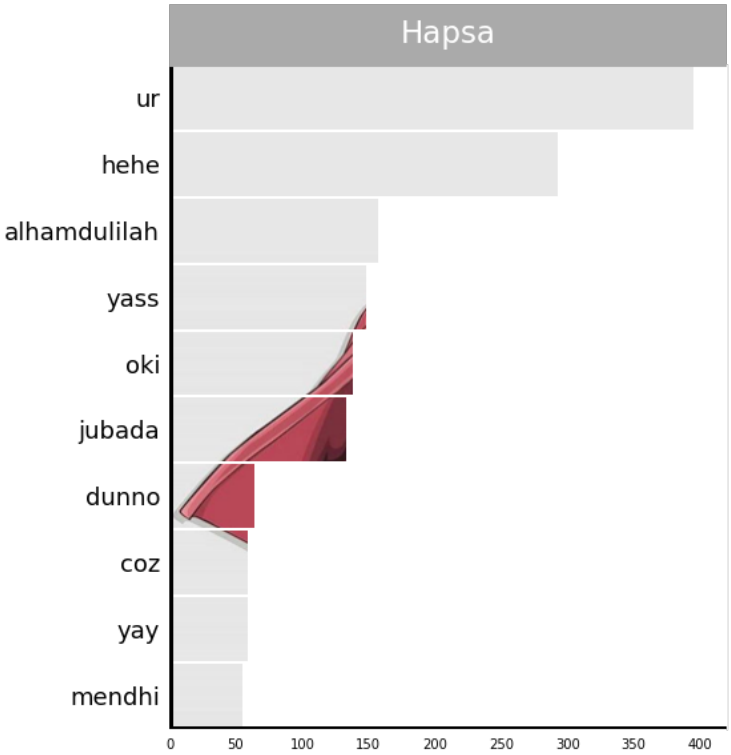


Figure 12: most common words used by user 2

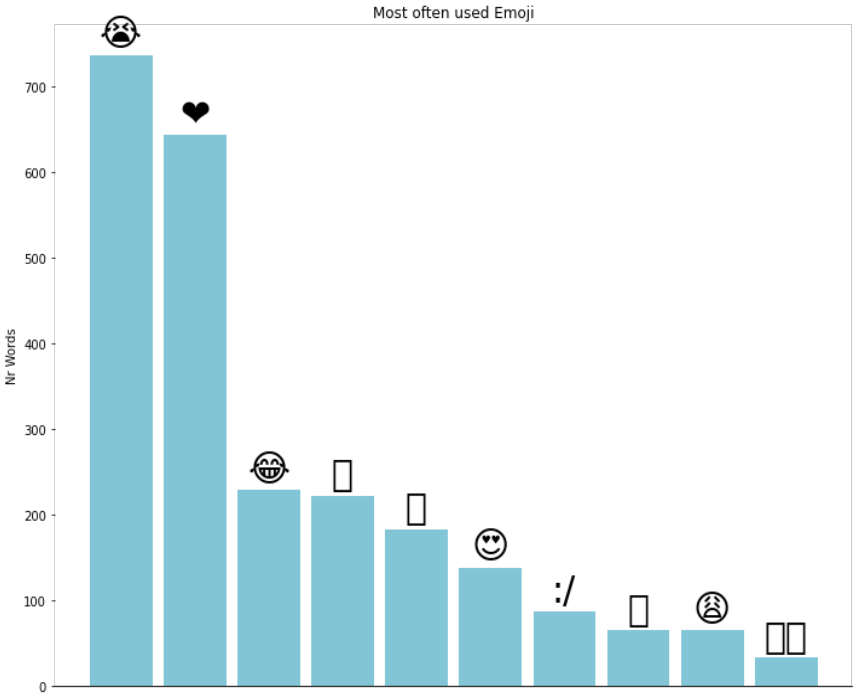


Figure 13: most used emoji by user 1

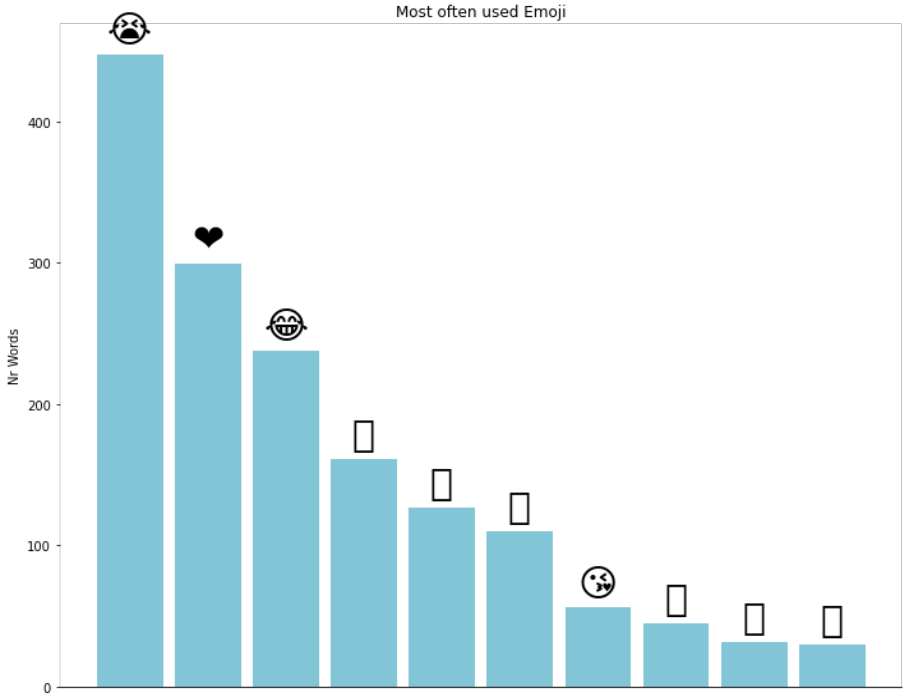


Figure 14: most used emoji by user 2