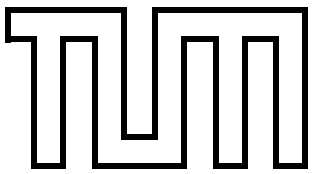


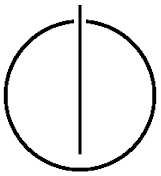
FAKULTÄT FÜR INFORMATIK

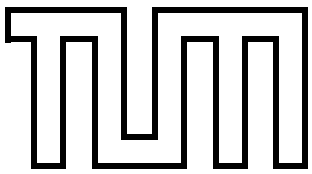
DER TECHNISCHEN UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Informatics

Analyzing Neurodegenerative Diseases with Web Chatbot Typing Behavior

Nicolas Othmar Theodarus





FAKULTÄT FÜR INFORMATIK

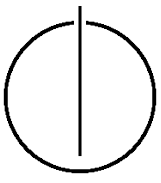
DER TECHNISCHEN UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Informatics

Analyzing Neurodegenerative Diseases with Web Chatbot Typing Behavior

Analyse neurodegenerativer Krankheiten anhand des Tippverhaltens von Web-Chatbots

Author:	Nicolas Othmar Theodarus
Supervisor:	Prof. Dr.-Ing. J org Ott
Advisor:	Yifan Yang, M.Sc.
Date:	03.12.2024



I assure the single handed composition of this bachelor thesis only supported by declared resources,

Munich, 03.12.2024 Nicolas Othmar Theodarus

Acknowledgements

Thank you everyone :D!

Abstract

Neurodegenerative diseases are chronic conditions that destroy and damage part of nervous system of the sufferer over time, especially the brain. This diseases pose a significant challenge for general public health, since the damages are permanent and incurable. This condition happens mainly on elderly people, given that aging is the greatest risk factor. Moreover, early detection of these diseases are inefficient, impractical and only have minuscule success percentage. There is a need for better detection methods that are cost-effective, user-friendly and accurate.

This thesis aims to pave the way of developing the aforementioned better detection methods. This thesis proposes a solution that involve developing a mobile optimized web application to gather typing data from users of different age groups. A clean and robust architecture structure is utilised to guarantee reliability, scalability and maintainability. It should also be ensured that the application is able to effectively process and save the collected data, so that the data can be used for research purposes. The application can also then be developed further with more advanced features. An example of such additional feature would be an analysis section, where typing behaviour data of a person can instantly be analysed with a click of a button.

An analysis of the data will be performed with the goal to find mathematical properties. This mathematical properties can then be used to categorize each user into their corresponding age groups. Understanding whether certain biomakers, e.g. typing pattern, can be used to differentiate characteristics of a person is the main focus of this thesis. The conclusion derived from this thesis could give insight into the feasibility of utilising biomarkers to effectively monitor health condition of the user. Specifically, the author hopes that this findings would be beneficials for research on detecting early signs of neurodegenerative disorders effectively with a simple method of collecting typing pattern data.

Zusammenfassung

Neurodegenerative Krankheiten sind chronische Erkrankungen, die im Laufe der Zeit Teile des Nervensystems der Betroffenen, insbesondere das Gehirn, zerstören und schädigen. Diese Krankheiten stellen eine große Herausforderung für die allgemeine öffentliche Gesundheit dar, da die Schäden dauerhaft und unheilbar sind. Sie treten vor allem bei älteren Menschen auf, da das Älterwerden der größte Risikofaktor ist. Darüber hinaus ist die Früherkennung dieser Krankheiten ineffizient, unpraktisch und hat nur einen verschwindend geringen Erfolgsanteil. Es besteht ein Bedarf an besseren Erkennungsmethoden, die kostengünstig, benutzerfreundlich und genau sind.

Ziel dieser Arbeit ist es, den Weg für die Entwicklung besserer Erkennungsmethoden zu ebnen. In dieser Arbeit wird eine Lösung vorgeschlagen, die die Entwicklung einer für Mobilgeräte optimierten Webanwendung beinhaltet, um Tippdaten von Benutzern verschiedener Altersgruppen zu sammeln. Es wird eine saubere und robuste Architekturstruktur verwendet, um Zuverlässigkeit, Skalierbarkeit und Wartbarkeit zu gewährleisten. Es sollte auch sichergestellt werden, dass die Anwendung in der Lage ist, die gesammelten Daten effektiv zu verarbeiten und zu speichern, so dass die Daten für Forschungszwecke verwendet werden können. Die Anwendung kann dann auch mit erweiterten Funktionen weiterentwickelt werden. Ein Beispiel für eine solche zusätzliche Funktion wäre ein Analysebereich, in dem die Daten zum Tippverhalten einer Person mit einem Klick auf eine Schaltfläche sofort analysiert werden können.

Die Analyse der Daten wird mit dem Ziel durchgeführt, mathematische Eigenschaften zu finden. Diese mathematischen Eigenschaften können dann verwendet werden, um jeden Nutzer in die entsprechenden Altersgruppen einzuteilen. Das Hauptaugenmerk dieser Arbeit liegt auf der Frage, ob bestimmte Biomacher, wie z.B. das Tippverhalten, zur Unterscheidung von Merkmalen einer Person verwendet werden können. Die aus dieser Arbeit abgeleiteten Schlussfolgerungen könnten Aufschluss darüber geben, inwieweit Biomarker zur effektiven Überwachung des Gesundheitszustands des Nutzers eingesetzt werden können. Insbesondere hofft der Autor, dass diese Erkenntnisse für die Forschung zur Erkennung früher Anzeichen von neurodegenerativen Erkrankungen mit einer einfachen Methode zur Erfassung von Tippmusterdaten von Nutzen sein könnten.

Contents

AD Alzheimer's Disease

PD Parkinson's Disease

Outline of the Thesis

Chapter 1: Introduction
Text

Chapter 2: Background
Text

Chapter 3: Related Work
Text

Chapter 4: Requirements Elicitation
Text

Chapter 5: System Design
Text

Chapter 6: Case Study/Evaluation
Text

Chapter 7: Summary
Text

Chapter 1

Introduction

Alzheimer's Disease (AD) and Parkinson's Disease (PD), are chronic and progressive neurodegenerative diseases that primarily affect the nervous system. This diseases lead to the degeneration of motor and cognitive abilities. These diseases are often irreversible and incurable, posing significant public health challenges. As populations age, the risk of such disorders increase substantially, making early detection crucial.

Historically, the diagnosis of neurodegenerative diseases has been done throuh clinical observations, imaging, and biomarkers. Even though medical technologies have improved significantly over the years, there are still many challenges for early diagnosis. One of the main reason for this is because neurodegenerative diseases develop gradually over time. Early symptoms of these diseases can easily be overlooked or mistaken as normal aging [VRW23]. As a result, these early symptons are often ignored until the disease itself has reached a critical stage, where the chance of treatment declines significantly. Furthermore, current diagnostic methods are expensive, invasive, and often inaccessible to a large portion of the population.

A better method is obviously needed to battle these insidious diseases. The rise of technologies such as smartphones open up new possibilities for early detection of these conditions. One example of such possibility is to utilise a phone application as a mean to detect early neurodegenerative diseases. Studies have shown that one of the effect of these diseases, e.g. impairments of motoric functions, will be reflected on how a person types [MFQM18]. This changes in typing behaviors, such as typing speed, accuracy, and keystroke patterns can then be recorded by the application. The data acquired by this application might be able to be used to analyse early symptons of AD and PD.

1.1 Problem

It is clear from the facts mentioned above, that neurodegenerative diseases are problems that need to be addressed. World Health Organisation estimates that there are approximately 50 million people worldwide affected by these disorders. Most of the sufferer are elderly, since age is one of the main risk factor. As the most common neurogedenerative disorder, AD still lacks an effective cure. It is even harder to treat the more progressive the disease progress. That is why the importance of an effective way to diagnose the disease early cannot be overstated. In the current state, however, misdiagnosis rates are still high, reaching up to 20%. Not only that, current diagnostic methods are either invasive, costly, or impractical for widespread use.

Similarly, PD, the second most common neurodegenerative disorder, has no cure and limited treatment options. For this disease too an effective mean for early diagnosis is of utmost importance. Since with a successful early detection, the progression of the disease can be slowed significantly, improving the quality of life of patients greatly. The current diagnostic methods for this disease, however, rely mostly on the observation of changes of motor symptoms. These methods, as previously discussed, are unreliable for many reasons.

In both conditions, early intervention can significantly improve patient outcomes, but existing diagnostic tools fail to provide a practical and accurate solution for early detection. There is a pressing need for non-invasive, cost-effective, and widely accessible methods to detect early signs of neurodegenerative diseases before the onset of significant symptoms.

1.2 Motivation

The motivation for this thesis comes from the urgent need to find better diagnostic methods for neurodegenerative diseases. Finding methods to effectively detect early these diseases would significantly improve general public health. Early detection allows for earlier interventions, which will then slow the progression of the diseases. This would improve treatment outcomes, and ultimately reduce the burden on healthcare systems.

From a scientific point of view, the research on using digital biomarker, i.e. typing pattern, as a mean to detect neurodegenerative diseases is underexplored. This research could give insights into how effective common tools and activities can be used to improve public health. Typing has became a common daily activity for most modern human, especially with the widespread use of smartphones and chat applications. This means that it can

be a low-cost and non-invasive method to detect subtle motor or cognitive impairments. In the most ideal case, the subject would not even notice that they are being monitored for these diseases and can go on about enjoying their daily lives.

Taking advantage of these common daily activities could help reduce the risk of neurodegenerative diseases, by diagnosing them as early as possible. Furthermore, these methods would also be accessible to people that lives in less developed countries with less developed medical technologies. This would ensure equal chances to fight against these neurodegenerative diseases. By developing a mobile-optimized web application that can collect and analyze typing data in real time, it would also become easier to monitor people's condition. Especially the condition of those that are more prone to this diseasesm, i.e. the elderly.

1.3 Objectives

Developing a web-based chat application that can capture and analyze typing behavior for early detection of neurodegenerative diseases is an ambitious goal. That goal is unfortunately not in the scope of this thesis. The primary objective of this thesis is to explore whether it is possible to determine the age of a user based on their typing patterns. The author believe that this thesis will pave the way for a more advanced research on this matter. This thesis wants to show that typing pattern can be used to identify the characteristics of the user, in this case, the age group.

Specifically, this thesis aims to:

1. Develop mobile-optimized chat application that collects the users' typing data. The main focus is to collect samples from individuals across different age groups.

2. Practice clean architecture and secure coding practices to make sure the application is reliable, scalable, and maintainable. The chat application will be designed to be user-friendly for all age groups, specifically the elderly.

3. Analyze the gathered typing behavior data to identify patterns or statistical distributions that may correlate with the user's age. Metrics such as typing speed, keystroke intervals, and error rates will be examined to determine if they can give indication of the user's age group.

4. Evaluate the accuracy of using typing behavior as a predictor of age.

The identified patterns need to be consistent enough to be able to reliably be used to estimate the user's age group.

The goal of this thesis is to research the feasibility of using data of typing behavior gathered by the application to profile the user in an age group. If this is achieved, the author hopes that this could give insights that would be valuable for future research in user profiling or cognitive assessments. The application could also be further developed to be able to analyse more complex matters, such as early signs of neurodegenerative disorders. Another possible improvement would be adding real-time analysis of the typing pattern and integration with healthcare systems. This would be beneficials for patients and clinicians alike.

Chapter 2

Background

Note: Describe each proven technology / concept shortly that is important to understand your thesis. Point out why it is interesting for your thesis. Make sure to incorporate references to important literature here.

Chapter 3

Related Work

Note: Describe related work regarding your topic and emphasize your (scientific) contribution in ***contrast*** *to existing approaches / concepts / workflows. Related work is usually current research by others and you defend yourself against the statement: "Why is your thesis relevant? The problem was already solved by XYZ." If you have multiple related works, use subsections to separate them.*

Chapter 4

Requirements Elicitation

Note: This chapter follows the Requirements Analysis Document Template in [BD09]. ***Important:*** *Make sure that the whole chapter is independent of the chosen technology and development platform. The idea is that you illustrate concepts, taxonomies and relationships of the application domain independent of the solution domain! Cite [BD09] several times in this chapter.*

4.1 Overview

Note: Provide a short overview about the purpose, scope, objectives and success criteria of the system that you like to develop.

4.2 Current System

Note: This section is only required if the proposed system (i.e. the system that you develop in the thesis) should replace an existing system.

4.3 Proposed System

Note: If you leave out the section "Current system", you can rename this section into "Requirements".

4.3.1 Functional Requirements

Note: List and describe all functional requirements of your system. Also mention requirements that you were not able to realize. The short title should be in the form "verb objective"

FR1 **Short Title**: Short Description.

FR2 **Short Title**: Short Description.

FR3 **Short Title**: Short Description.

4.3.2 Nonfunctional Requirements

Note: List and describe all nonfunctional requirements of your system. Also mention requirements that you were not able to realize. Categorize them using the FURPS+ model described in [BD09] without the category ***functionality*** *that was already covered with the functional requirements.*

NFR1 **Category**: Short Description.

NFR2 **Category**: Short Description.

NFR3 **Category**: Short Description.

4.4 System Models

Note: This section includes important system models for the requirements analysis.

4.4.1 Scenarios

Note: If you do not distinguish between visionary and demo scenarios, you can remove the two subsubsections below and list all scenarios here.

Visionary Scenarios

Note: Describe 1-2 visionary scenario here, i.e. a scenario that would perfectly solve your problem, even if it might not be realizable. use our scenario description template in form of a table.

Demo Scenarios

Note: Describe 1-2 demo scenario here, i.e. a scenario that you can implement and demonstrate until the end of your thesis. use our scenario description template in form of a table.

4.4.2 Use Case Model

Note: This subsection should contain a UML Use Case Diagram including roles and their use cases. You can use colors to indicate priorities. Think about splitting the diagram into multiple ones if you have more than 10 use cases. ***Important:*** *Make sure to describe the most important use cases using the use case table template. Also describe the rationale of the use case model, i.e. why you modeled it like you show it in the diagram.*

4.4.3 Analysis Object Model

Note: This subsection should contain a UML Class Diagram showing the most important objects, attributes, methods and relations of your application domain including taxonomies using specification inheritance (see [BD09]). Do not insert objects, attributes or methods of the solution domain. ***Important:*** *Make sure to describe the analysis object model thoroughly in the text so that readers are able to understand the diagram. Also write about the rationale how and why you modeled the concepts like this.*

4.4.4 Dynamic Model

*Note: This subsection should contain dynamic UML diagrams. These can be a UML state diagrams, UML communication diagrams or UML activity diagrams.****Important:*** *Make sure to describe the diagram and its rationale in the text.* ***Do not use UML sequence diagrams.***

4.4.5 User Interface

Note: Show mockups of the user interface of the software you develop and their connections / transitions. You can also create a storyboard. ***Important:*** *Describe the mockups and their rationale in the text.*

Chapter 5

System Design

Note: This chapter follows the System Design Document Template in [BD09]. You describe in this chapter how you map the concepts of the application domain to the solution domain. Some sections are optional, if they do not apply to your problem. Cite [BD09] several times in this chapter.

5.1 Overview

Note: Provide a brief overview of the software architecture and references to other chapters (e.g. requirements analysis), references to existing systems, constraints impacting the software architecture.

5.2 Design Goals

Note: Derive design goals from your nonfunctional requirements, prioritize them (as they might conflict with each other) and describe the rationale of your prioritization. Any trade-offs between design goals (e.g., build vs. buy, memory space vs. response time), and the rationale behind the specific solution should be described in this section

5.3 Subsystem Decomposition

Note: Describe the architecture of your system by decomposing it into subsystems and the services provided by each subsystem. Use UML class diagrams including packages / components for each subsystem.

5.4 Hardware Software Mapping

Note: This section describes how the subsystems are mapped onto existing hardware and software components. The description is accompanied by a UML deployment diagram. The existing components are often off-the-shelf components. If the components are distributed on different nodes, the network infrastructure and the protocols are also described.

5.5 Persistent Data Management

Note: Optional section that describes how data is saved over the lifetime of the system and which data. Usually this is either done by saving data in structured files or in databases. If this is applicable for the thesis, describe the approach for persisting data here and show a UML class diagram how the entity objects are mapped to persistent storage. It contains a rationale of the selected storage scheme, file system or database, a description of the selected database and database administration issues.

5.6 Access Control

Note: Optional section describing the access control and security issues based on the nonfunctional requirements in the requirements analysis. It also describes the implementation of the access matrix based on capabilities or access control lists, the selection of authentication mechanisms and the use of encryption algorithms.

5.7 Global Software Control

Note: Optional section describing describing the control flow of the system, in particular, whether a monolithic, event-driven control flow or concurrent processes have been selected, how requests are initiated and specific synchronization issues

5.8 Boundary Conditions

Note: Optional section describing the use cases how to start up the separate components of the system, how to shut them down, and what to do if a component or the system fails.

Chapter 6

Case Study / Evaluation

Note: If you did an evaluation / case study, describe it here.

6.1 Design

Note: Describe the design / methodology of the evaluation and why you did it like that. E.g. what kind of evaluation have you done (e.g. questionnaire, personal interviews, simulation, quantitative analysis of metrics, what kind of participants, what kind of questions, what was the procedure?

6.2 Objectives

Note: Derive concrete objectives / hypotheses for this evaluation from the general ones in the introduction.

6.3 Results

Note: Summarize the most interesting results of your evaluation (without interpretation). Additional results can be put into the appendix.

6.4 Findings

Note: Interpret the results and conclude interesting findings

6.5 Discussion

Note: Discuss the findings in more detail and also review possible disadvantages that you found

6.6 Limitations

Note: Describe limitations and threats to validity of your evaluation, e.g. reliability, generalizability, selection bias, researcher bias

Chapter 7

Summary

Note: This chapter includes the status of your thesis, a conclusion and an outlook about future work.

7.1 Status

Note: Describe honestly the achieved goals (e.g. the well implemented and tested use cases) and the open goals here. if you only have achieved goals, you did something wrong in your analysis.

7.1.1 Realized Goals

Note: Summarize the achieved goals by repeating the realized requirements or use cases stating how you realized them.

7.1.2 Open Goals

Note: Summarize the open goals by repeating the open requirements or use cases and explaining why you were not able to achieve them. ***Important:*** *It might be suspicious, if you do not have open goals. This usually indicates that you did not thoroughly analyze your problems.*

7.2 Conclusion

Note: Recap shortly which problem you solved in your thesis and discuss your ***contributions*** *here.*

7.3 Future Work

Note: Tell us the next steps (that you would do if you have more time. be creative, visionary and open-minded here.

Appendix A

e.g. Questionnaire

Note: If you have large models, additional evaluation data like questionnaires or non summarized results, put them into the appendix.

Appendix B

Tips for writing a thesis in TeX

B.1 using this template

This template tries to achieve a separation of the template itself and the parts that are specific to the thesis. Ideally, the template does not have to be edited.

The content of the thesis shall be added to the following files and folders:

- the `.tex`-files in the `chapters`-folder shall contain the description of your scientific work.
- the `.tex`-files in the `resources`-folder contain templates and examples, into which metadata, settings and organisational information about the thesis can be entered.
- the `thesis.bib`-file shall contain a list of the literature, that you cite in your thesis.

B.2 General tips

Track your work on this thesis with a version control system such as git.

In your TeX source code, use one line per sentence. This facilitates spotting excessively long sentences. Also, it makes the tracking of changes by the version control system more useful. If you add line breaks after a fixed number of columns instead, a change affects all subsequent lines of the paragraph, even though the actual contend has not been changed.

It is recommended to create a folder, in which all images, that are included in this document are stored. See the `resources/settings.tex`-file, on how to add this folder to the default graphics path, so only the filenames have to be entered, when including an image.

List of Figures

List of Tables

Bibliography

[BD09] Bernd Bruegge and Allen H Dutoit. *Object Oriented Software Engineering Using UML, Patterns, and Java*. Prentice Hall, 2009.

[MFQM18] Tara L. McIsaac, E. Nora Fritz, Lori Quinn, and Lisa M. Muratori. Cognitive-motor interference in neurodegenerative disease: A narrative review and implications for clinical management. *Frontiers in Psychology*, 9(02061), 2018.

[VRW23] Manera Valerias, Erika Rovini, and Peter Wais. Early detection of neurodegenerative disorders using behavioral markers and new technologies: New methods and perspectives. *Frontiers in Aging Neuroscience*, 15(1149886), 2023.