

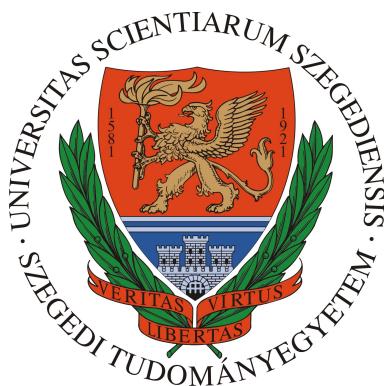
Telemedicine Engineering: Modeling and Analyzing Data Quality in Telemedicine Systems and the Impact of Novel Design Patterns on Productivity in Modern Web Application Development

SUMMARY OF THE PHD THESIS

Zoltán Richárd Jánki

Supervisor:
Vilmos Bilicki, PhD
assistant professor

Doctoral School of Computer Science
Department of Software Engineering
Faculty of Science and Informatics
University of Szeged



Szeged
2023

1 Introduction

Telemedicine is a rapidly developing field supported by an increasing number of software applications. Today's telemedicine systems are rarely founded on a simple client-server architecture but rather on data paths that can be extremely complex as it is depicted in Figure 1. Today, microservice-level solutions and the integration of existing, well-functioning solutions are extremely prevalent. In the course of our experience, we have also encountered this reality, and it is crucial to place emphasis not only on the software that supports telemedicine but also on the tools and methods that facilitate the development of telemedicine software and are easily adaptable to software development processes in other fields.

The dissertation examines the primary challenges of telemedicine system development from the perspectives of data quality and development productivity. We present a variety of modeling, simulation, graph-based, and approximation-based techniques for measuring the data quality and consistency of complex telemedicine systems. Through the application of the Web Data Access Object (WebDAO) design pattern, my research presents a self-developed Software Development Kit (SDK) that not only improves data integrity but also increases productivity. In addition to measuring the required development time, the results are validated by comparing repository metrics. We found, based on my experiences, that structural and architectural design patterns play very important roles and have many positive effects. As a result, we devised various Natural Language Processing (NLP) techniques in order to provide an in-depth understanding of the applied best practices and make recommendations for contemporary Web development.

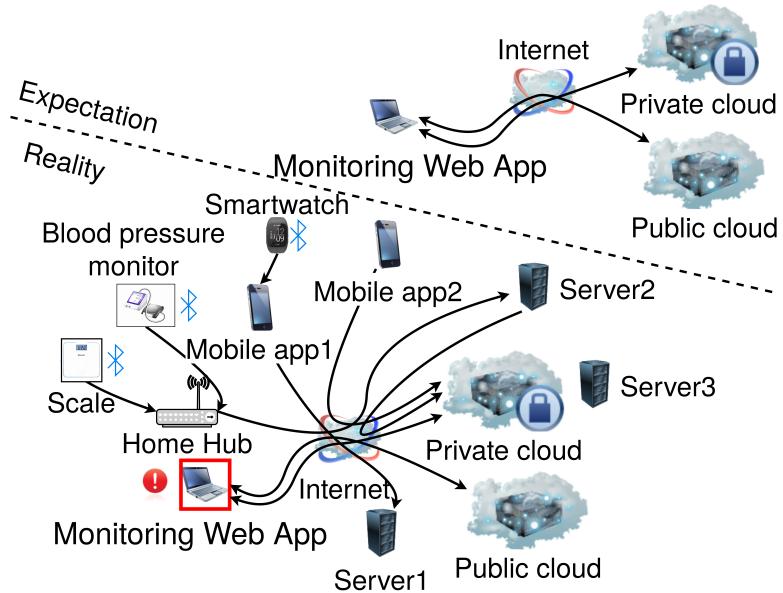


Figure 1: Expected structure of telemedicine systems versus data paths in real telehealth systems

2 Thesis Group I: Analysis of Data Consistency and Data Quality in Distributed Telemedicine Systems

I created a modeling and simulation environment for measuring the consistency and the data quality in distributed telemedicine systems. Simulation results showed that the latency can significantly influence the consistency and data quality. Based on the results, I proposed a taxonomy for managing the trade-off problem present in distributed telemedicine systems and gave formal specification of such a system. After further analysis, it was shown that a graph representation can be an efficient procedure to evaluate distributed systems. Taking the advantage of the DAG structure and different graph algorithms, the complexity of the problem can be simplified. Finally, I developed a dedicated framework called TSMOF that can analyze a focused part of the data quality related state space of a system under given circumstances by using emulation, simulation and approximation techniques.

Publications related to this thesis group: [J1], [J2], [J3], [FR2], [FR3]

A telemedicine system consists of a number of components that may lead to complex data paths. During the journey of the data, it goes through various processes consisting of transformations, aggregations and mappings that are potential steps in which data can be corrupted. Data corruption is often caused by latency or improper system configuration, so it is important to find the best matching settings to the use-case. This thesis group presents a taxonomy of telemedicine use-cases and a recommended system design and configuration. The classification is supported by the Consistency, Availability and Partition tolerance (CAP) and the PACELC Theorems. The taxonomy is validated with data quality simulations based on a formal specification of such a telemedicine system. Here, we introduce graph-based approach to evaluate the simulation results efficiently. However, the data quality related state space of such a system can be extremely large, so it is reasonable to analyze a more focused part of it. Finally, we show an approximation technique to reduce the complexity of the data quality related state space and provide precise results for the data quality in a real-time telemedicine system based on real executions.

2.1 Thesis I/1: Taxonomy of Distributed Telemedicine Systems Based on the CAP and PACELC Theorems

I introduced the trade-off problem in distributed telemedicine systems in terms of CAP Theorem. I presented a formal method for modeling and simulating the operation of such systems. I identified the presence of the trade-off problem with the help of telemedicine use-cases and measured that consistency and data quality is influenced by latency. I proposed a taxonomy for distributed telemedicine systems taking into account the trade-off between consistency and availability.

Publications related to this thesis: [J1], [FR2], [FR3]

Based on Erich Brewer's CAP Theorem [3], there is no distributed system that can guarantee at most two of the three desirable properties: consistency (C), availability (A) and partition tolerance (P). Due to this limitation a trade-off has to be made and this

observation formed the extension of CAP called PACELC Theorem [1]. PACELC states that in the case of network partitioning (P), a trade-off has to be made between availability (A) and consistency (C), but else (E), when the system is running normally in the absence of partitions, another trade-off has to be made between latency (L) and consistency (C). As of the diversity of telemedicine, it is crucial to find the proper balance of capabilities based on the use-case. To find the balance, we examined the data considering the importance of consistency. Using the idea presented by Peter Bailis et al. [2], we can assign a k -staleness parameter to the data that describes we accept data older than k -versions at most. After modeling a parameterizable system, the simulations showed that if k is set to 0, we can almost get 100% of consistency, however latency can cause a drastic decline. Consistency is reducing by 5% after increasing k by 1, and latency can decrease consistency by another 5% after increasing the latency by 100 milliseconds. Based on the results, we constructed the taxonomy for the trade-off problem in distributed telemedicine systems that can be seen in Table 1. We categorized the use-cases into 3 classes: non-offline telemedicine, semi-offline telemedicine and offline telemedicine taking into account whether caching is permitted or not. We have seen that data staleness may occur independently from caching, so we further separated the use-cases based on the possibility of data staleness. If caching is not permitted, it is recommended to use PC/EL system. If cache is deployed, it is advised to focus on consistency but with a low k -staleness parameter. In the case of offline telemedicine, availability is more important than consistency, and a higher k -staleness parameter is suggested.

Table 1: Telemedicine system trade-off taxonomy

Use-case category	Possibility of staleness	Recommendation for system selection
Non-offline telemedicine	No	PC/EL without cache
	Yes	PC/EC without cache
Semi-offline telemedicine	No	PC/EL with cache higher k -staleness parameter
	Yes	PC/EL with cache lower k -staleness parameter
Offline telemedicine	No	PA/EL with cache higher k -staleness parameter
	Yes	PA/EL with cache higher k -staleness parameter

2.2 Thesis I/2: Graph-based Method for Analyzing Data Quality in Telemedicine Systems

I showed - using the formal method defined in thesis I/1 - that the data quality related state space of a distributed system is described with graph that have a DAG structure. Taking the advantage of the structure, I made graph clustering on the weakly connected components of the graph and optimized the complexity of the data quality related state space.

Publications related to this thesis: [J2]

Simulation is widely accepted and utilized technique for analyzing how a system would perform under given circumstances [12]. L. Lamport introduced the Temporal Logic of Actions (TLA) for specifying concurrent systems [9]. After gaining experience in TLA, we modeled several telemedicine use-cases and scenarios and examined how consistency and data quality may change due to caching and latency. The system model elaborated in TLA can be verified using the TLA+ Model Checker called TLC [14]. The graph output produced by TLC contains all possible states that a system can have. A relatively small system can have millions of nodes and edges, so it is difficult to evaluate the entire graph in one piece. After analyzing the structure of the graphs, it turned out they contain cycles, however only the leaves have self-loop edges. After removing the terminating edges, the graphs have a Directed Acyclic Graph (DAG) structure. Knowing that we have DAG structure, the analysis of the simulations can be significantly simplified by using graph algorithms. In the graph, each simulation preformed by TLC under given circumstances is represented by a weakly connected component. Here, we performed a clustering of the weakly connected components based on the data quality values found in the leaves. It is found that if components have the same data quality values, they have the same graph structure. Table 2 shows that with clustering, only 2% of the data quality related state space needs to be examined. However, it is crucial to observe that some clusters contain only one or two components, while others may contain thousands. Taking the advantage of the DAG structure, graph algorithms can be executed in linear time. All in all, the complexity of the data quality related state space can be significantly reduced by using a graph-based approach and a clustering technique together.

Table 2: Number of weakly connected components and generated clusters with different k -staleness parameter values

k -staleness parameter	Number of components constructed via model checking	Number of clusters obtained from clustering method based on data quality
0	276,862	5,383
1	288,988	4,735
2	282,104	4,112
3	282,032	4,139

2.3 Thesis I/3: Modeling, Emulation, and Simulation of Distributed Telemedicine Systems Employing Approximation Techniques

I proposed approximation techniques to find the best fitting PDFs that can describe the processes of a complex system. I designed a framework that can examine a focused part of the data quality related state space of a system under given circumstances. I demonstrated the effectiveness of the framework with executions of real algorithms. I showed that a significant data quality loss can be observed due to latency in real-time telemedicine systems.

Publications related to this thesis: [J3]

In telemedicine, data path can be complex including multiple data manipulation and aggregation phases. It is seen that latency can significantly influence the data quality that is more crucial in the case of real-time systems because buffering and delayed processing are not permitted. Our real-time telemedicine use-case consists of four interdependent processes including electrocardiogram (ECG) measurements, data upload, atrial fibrillation (AF) detection and AF classification phases. If processes take so long, system cannot remain real-time and valuable data portions must be dropped to maintain the real-time capability but it leads to data quality loss. To observe the amount of data quality reduction, modeling and simulation techniques can be used. We modeled a real-time ECG system including AF detection and AF classification algorithms, it is showed in Code [1]. Although such a system can be logically modeled, it is seen that the data quality related state space can be immense in the case of a minor system, and the evaluation of it would require years.

Code 1: Formal model of the latency in a real-time telemedicine system

```

CONSTANT MAX_NUM_WIN, WINDOW_SIZE, TIMEOUT_UP
VARIABLES numWin, now1, now2, now3, now4, dropList, ui, pi1, pi2

Step(t) == /\ IF numWin[t] < MAX_NUM_WIN
    THEN /\ numWin' = [numWin EXCEPT ![t] = numWin[t] + 1]
        /\ now1' = [now1 EXCEPT ![t] = now1[t] + WINDOW_SIZE]
        /\ now2' = [now2 EXCEPT ![t] = now1'[t] + ui[t]]
        /\ IF now3[t] < now2'[t] + TIMEOUT_UP
            THEN /\ dropList' = [dropList EXCEPT ![t] = <<FALSE>> \o
                dropList[t]]
                /\ now3' = [now3 EXCEPT ![t] = now2'[t] + pi1[t]]
                /\ IF now3'[t] > now4[t]
                    THEN /\ now4' = [now4 EXCEPT ![t] = now3'[t] +
                        pi2[t]]
                        ELSE /\ UNCHANGED <<now4>>
                ELSE /\ dropList' = [dropList EXCEPT ![t] = <<TRUE>> \o
                    dropList[t]]
                    /\ UNCHANGED <<now3, now4>>
            /\ ui' = [ui EXCEPT ![t] = (UPLOAD_LATENCY[numOp[t] + 1])]
            /\ pi1' = [pi1 EXCEPT ![t] = (PROC_TIME_ALG_1[numOp[t] +
                1])]
            /\ pi2' = [pi2 EXCEPT ![t] = (PROC_TIME_ALG_2[numOp[t] +
                1])]
        ELSE /\ UNCHANGED vars

```

If an implementation of the system or a process is available, we can emulate its behaviour with virtualization, and multiple variation of configurations can be tested. Based on the emulation results, we can see how much latency is generated by a given process or the system. Using the emulation results, we can construct the probability density function (PDF) of the latency with approximation techniques. Using the PDF, we can generate random values from the distribution that describes the process. If all critical components and processes can be emulated or approximated such a way, an accurate data quality can

be calculated for the entire system using the model and evaluating a focused part of the data quality related state space. Considering the selected use-case, the simulation results showed that data quality loss can be 4.25% after the first aggregation but after the second aggregation the reduction can be more than 20%.

3 Thesis Group II: A Productivity-enhancing Software Development Kit with Support for Hybrid Cloud

I introduced a novel design pattern called WebDAO and presented its significance in modern front-end application development. I showed that FHIR is a widely accepted standard in telemedicine projects and I created an implementation of it in the form of a SDK that supports hybrid cloud. I presented that my standardized installable SDK - that applies WebDAO, - improves the development productivity. By extending the analysis to open-source projects, repository metrics showed that WebDAO improves development productivity.

Publications related to this thesis group: [J4], [J6], [C1], [FR1], [FR4]

In software development, we often face with recurring or similar problems and challenges. Solutions for frequently returning issues will emerge best practices and the generalization of the best practices create design patterns [15]. The design patterns can not only solve problems in general but they have several good impacts on development processes. Besides code reusability and maintainability [6], productivity is not the last consideration. In this thesis group, we present an SDK that not only standardizes the front-end and back-end communication in modern Web development and supports hybrid cloud approaches but significantly increases development productivity by implementing the WebDAO design pattern. Finally, we present the importance of WebDAO by analyzing thousands of open-source Angular projects.

3.1 Thesis II/1: Standardized Front-end and Back-end Connection with WebDAO

I presented that the front-end and back-end communication can be standardized with the WebDAO design pattern in modern Web development. Especially in telemedicine, the FHIR standard makes a perfect combination with WebDAO. I created usage statistics about FHIR in Web development. I showed that the layered architecture changed due to the aspiration of serverless development and demonstrated how Angular application development utilizes the new architecture.

Publications related to this thesis: [J4], [J6]

In software development, it is crucial to design the software at a system level and define the architecture of it including the responsibilities of the components. In contemporary Web development, the so-called layered or n-tier architecture is one of the most widely utilized architectural patterns and a de facto standard as well [11]. The traditional layered architecture consists of five layers - from which one is optional - with well-defined responsibilities. Today, as serverless development gains more and more space, so the back-end

is getting thinner and more responsibilities are moved to the front-end. In Angular application development, the services implement the service layer of the traditional layered architecture but this approach segregates the top three layers from the two at the bottom. To maintain the communication between the front-end and the back-end standardized, it is recommended to introduce a widely accepted standard - such as Fast Healthcare Interoperability Resources (FHIR) [4]. This thesis point shows that FHIR is the most widely applied standard for healthcare systems and there are thousands of open-source projects that utilize FHIR all over the world. Besides the standardized domain model, it is proposed to move the Data Access Object (DAO) capabilities to the front-end. As it is displayed in Figure 2, we refer to this design pattern as WebDAO in the new layered architecture. During the development of telemedicine projects, we found that WebDAO plays an important role because it not only standardizes the communication between the front-end and the back-end, but several supplementary services can be added to our projects.

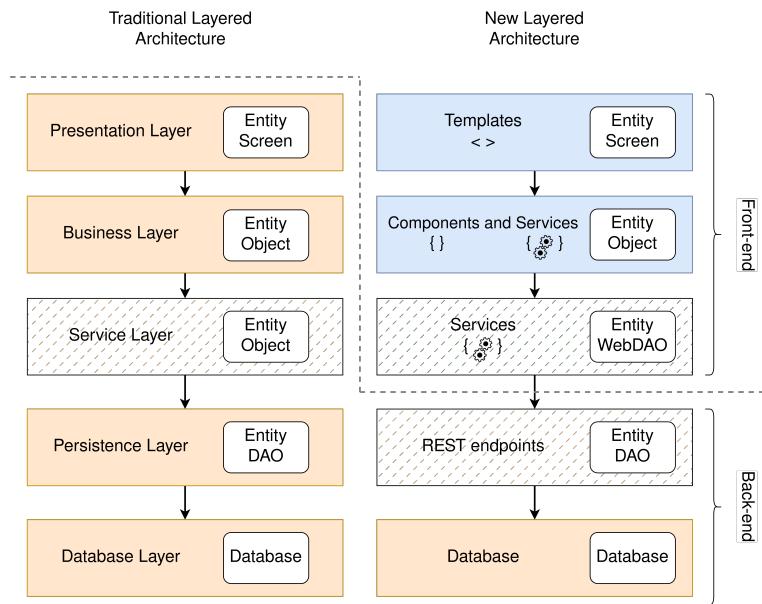


Figure 2: Traditional and modern layered architecture

3.2 Thesis II/2: The Application of WebDAO and Productivity Measurements in the Development of Telemedicine Application

I developed a SDK for telemedicine application development that supports hybrid cloud approach and applies the WebDAO design pattern. I proposed a methodology that shows the quantitative extent of development productivity improvement utilizing WebDAO. I made evaluations on the productivity involving real developers and showed that WebDAO can significantly improve development productivity.

Publications related to this thesis: [J6], [C1], [FR1], [FR4]

The application of the design patterns may have various effects on the software development processes. We developed the Inclouded SDK [7] that incorporates the WebDAO

capabilities, hybrid cloud support, polyglot persistence and standardized domain model using FHIR. Our SDK performed well during the development of a number of telemedicine pilots. It is said that DAO improves the development productivity due to code reusability and simplicity of the implementation. Since WebDAO has similar capabilities but it has been moved to the front-end, our hypothesis was that it improves the productivity as well. A further research question is that if WebDAO has positive effect on productivity, then what is the quantitative extent of this effect? To demonstrate the impacts of WebDAO, we involved university students studying computer science and measured their productivity. They had to implement Angular services for the four most important FHIR resources in our telemedicine pilot projects with and without Inclouded SDK. Inclouded SDK contains the required create, read, update and delete (CRUD) operations of the FHIR resources, the recommended further filter capabilities of the resources and the necessary FHIR interfaces. In Figure 3, it is seen that the development time is reduced by at least 50% if WebDAO is applied in Angular service development in the form of Inclouded SDK. Although, the number of involved students is quite small, the positive effect of DAO and WebDAO on productivity is quantitatively shown.

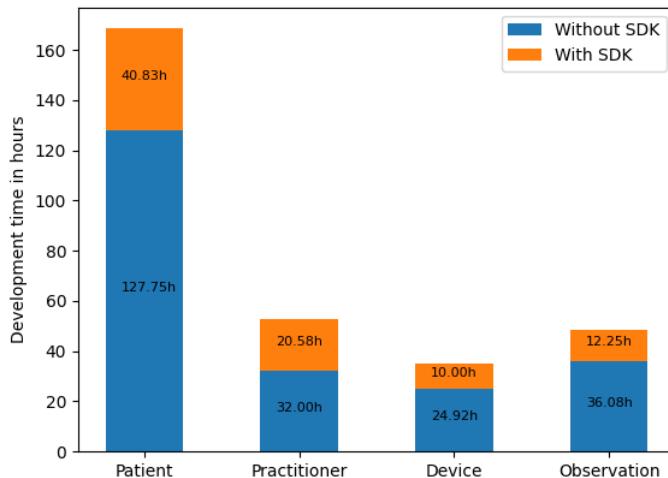


Figure 3: Average development time measured with and without the SDK

3.3 Thesis II/3: Productivity Measurement Based on Repository Metrics and the Application of WebDAO

I showed a productivity measurement method using repository metrics. Based on the results, I demonstrated that if WebDAO is applied in Angular application development, it can reduce the source code changes and the solutions for the issues require less time.

Publications related to this thesis: [J4]

Since the WebDAO design pattern has never been introduced before but we consider practically significant, we assumed that it is widely used in modern Web development, especially in Angular projects. If the assumption is correct, we can measure the produc-

tivity from other aspects as well [8, 10, 13]. To validate the assumption, we downloaded 19,116 open-source Angular projects and analyzed the implementations using the methods presented in thesis group III. and thesis III/1. We found that most of the developers use services for back-end communications which is a best practice, and many of the services are WebDAO. We compared the productivity of the projects based on the repository metrics. In Figure 4, it is seen that the number of commits and the number of issues are similar in both cases. The modifications in the source code is two times less if WebDAO is applied which indicates that these projects are designed more carefully and the waterfall model-like development seems to be less typical. Although the number of issues are quite similar, the average length and the standard deviation of them have significantly lower values which allows us to conclude that solving the issues require less time and the problem solving is more focused and more efficient. In conclusion, WebDAO is implicitly used in Angular application development and has a positive effect on productivity taking into account the repository metrics.

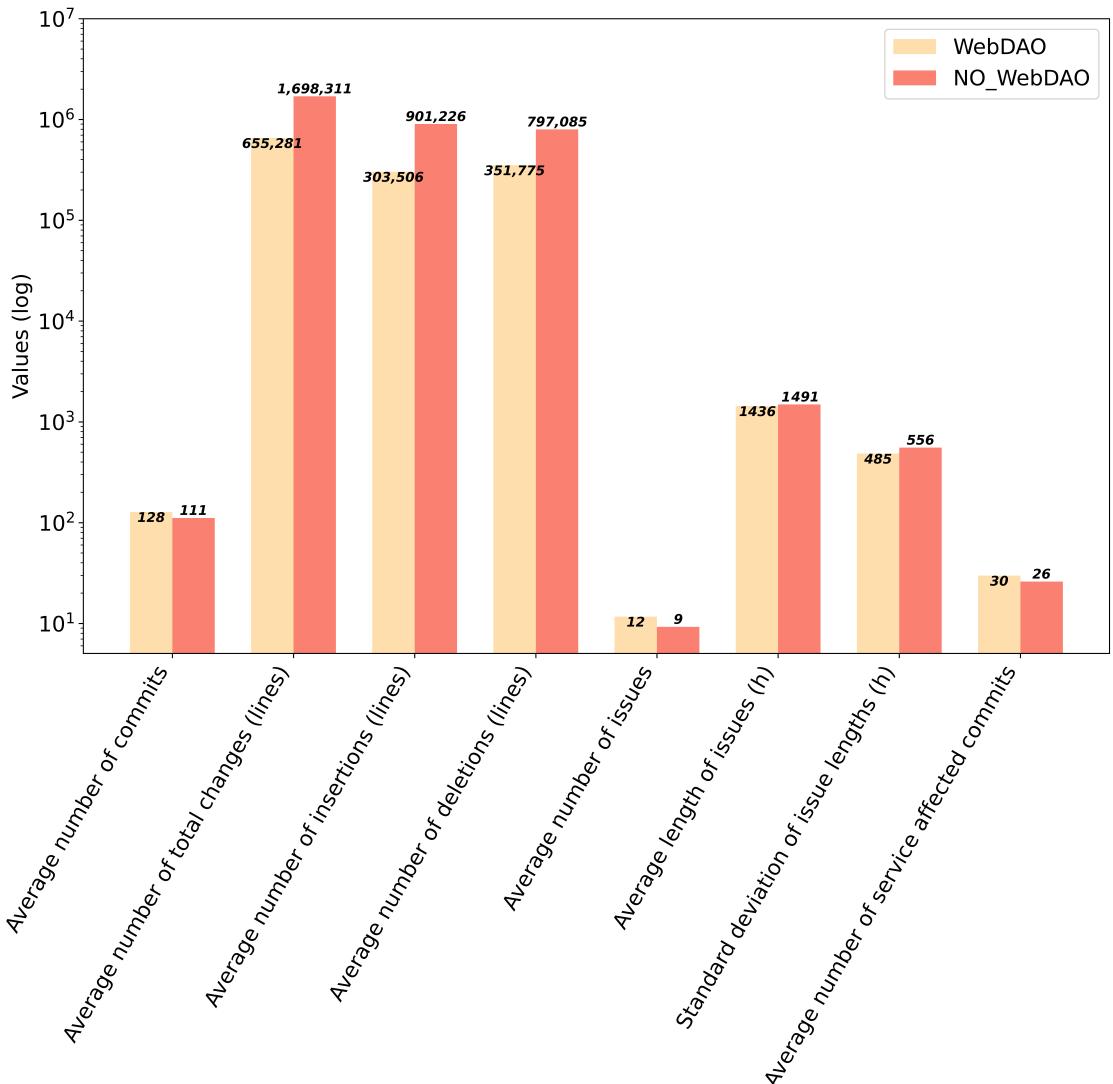


Figure 4: Project metrics dependent on whether or not WebDAO is used

4 Thesis Group III: Identifying Design Patterns Using Natural Language Processing Techniques

After introducing the WebDAO design pattern, I developed a Random Forest classifier and a BERT model to identify the design pattern in the open-source projects. I showed that WebDAO is often used implicitly in Angular projects. I observed that multiple Model-View-based design patterns are applied in modern Web development, so I elaborated a taxonomy of them and proposed use-cases for them. Finally, I implemented a ruleset that can be used to identify MVW design patterns with high accuracy in Angular projects using LLMs.

Publications related to this thesis group: [J4], [J5]

Design patterns are essential in software development because they offer generally applicable best practices with the use of which developers can create more maintainable, scalable and comprehensible software. Today, various architectural, structural and behavioral design patterns are known [5, 6] and their principles are already well-defined but it is seen that their applications, structures and roles may differ based on the used technology and framework. Referring to Chapter 3 and Chapter 4, we analyzed the softwares on a system level and presented architectural novelties. Based on the achieved results, we believe that it is crucial to spread the word about the importance of design patterns. More and more researchers deal with the identification and the automatic detection of design patterns with distinct methods based on different features. ML and especially NLP methods are the most popular and most powerful solutions today. In this thesis group, we present our NLP approaches to detect the WebDAO design pattern in Angular applications with a high accuracy and provide usage statistics as well. Our gained experiences in Angular application development motivated us to analyze the architectural solutions utilized in the projects. We found that a number of Model-View-based architectural design patterns can be implemented and can have positive practical impact. To clarify the differences and provide recommendations, we elaborated a taxonomy of MVW design patterns used in modern Web development. Often it is difficult to clearly identify which design pattern is used, so we intended to present a novel prompt engineering approach for detecting design patterns with LLMs.

4.1 Thesis III/1: Identification of the WebDAO Design Pattern with state-of-the-art AI techniques

I developed and compared two NLP methods for classifying Angular projects and services based on the presence of the WebDAO design pattern. I showed that WebDAO is an implicitly used design pattern in Angular application development and revealed that there is a number of projects that do not use services at all.

Publications related to this thesis: [J4]

Design patterns detection help better understanding the software and can greatly assist in the maintenance of systems. In Chapter 4, we showed the importance of WebDAO and that they are often present in Angular services. Here, we present our WebDAO classifier

that can point out Angular services that do not conform the design pattern and highlight the components in which code quality could be improved. Since the decision can be made based on the evaluation of a single file, we analyzed the services of Angular projects. We assumed that NLP methods can be very effective in this classification problem. We labeled 504 Angular services from which 247 was implementing WebDAO, and 257 did not apply the WebDAO design pattern. The services were selected from 112 mature open-source projects downloaded from GitHub. Firstly, we trained a Random Forest classifier that achieved an F1-score of 89.8%, along with a precision of 86.3% and a recall of 93.6%. Secondly, we implemented a Bidirectional Encoder Representations from Transformers (BERT) model to further improve the accuracy. With our BERT classifier, we achieved 92% of F1-score for both WebDAO and NO_WebDAO classes. With this trustworthy model, we analyzed all downloaded Angular project to provide statistics about the use of WebDAO. We found that the ratio of using WebDAO is quite similar on file-level (see Figure 5a) and including projects that contain services (see Figure 5b). In total, WebDAO is present only in 28.09% of the Angular projects, however 27% of the projects classified as NO_WebDAO projects do not include services at all. If the examined set contains only projects that use services, then WebDAO is present in a bit more than 1/3 of the projects.

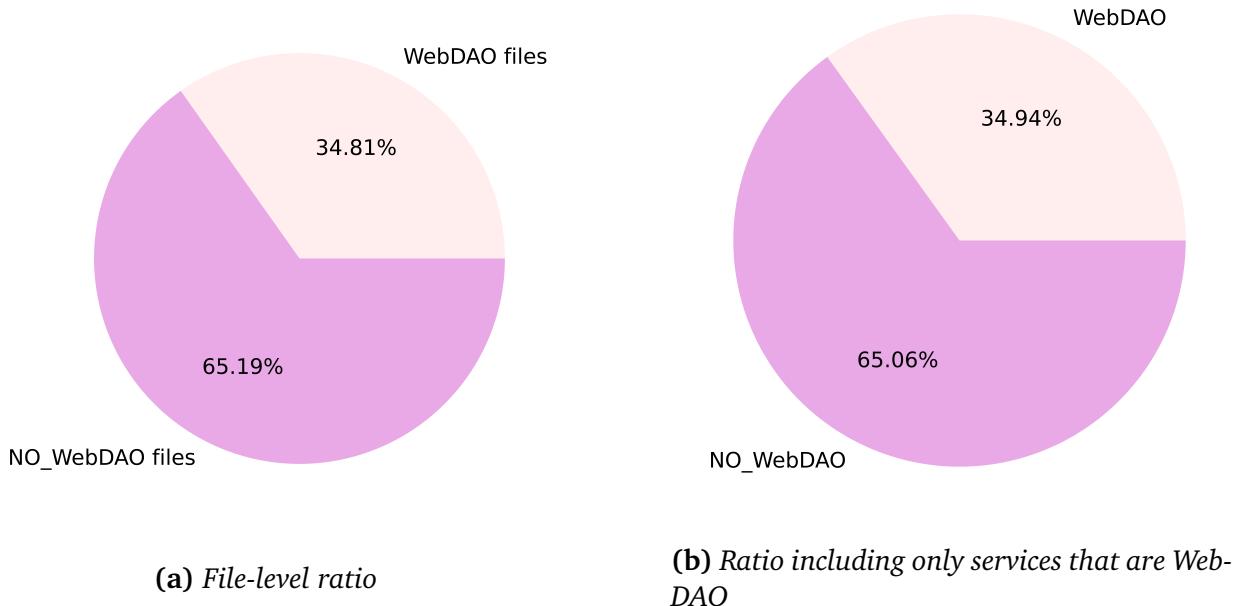


Figure 5: Ratio of projects utilizing WebDAO and ratio of Angular services that are WebDAO

4.2 Thesis III/2: Taxonomy of Model-View-Whatever Design Pattern Variants in Modern Web Development

I introduced the available variants of Model-View-based architectural design patterns in modern Web development. I proposed a taxonomy including formal definitions and recommendations when a specific MVW design pattern is appropriate. I gave Angular use-cases and examples for applying specific design patterns.

Publications related to this thesis: [J5]

There are a number of software design patterns that offer general best practices to have a maintainable and reusable source code. However, many variants of architectural design patterns have been created by the needs of different technologies and frameworks. Angular framework is often considered as a Model-View-ViewModel (MVVM) framework based on the best practices but it has never been confirmed by Google. Also, we found that developers often make solutions that contradicts the principles of MVVM. Here, we collect the MVW design pattern variants available in Angular development that can be easily adapted to other modern frameworks as well. Firstly, we categorized and depicted in Figure 7 the available design patterns based on which family they belong to. Secondly, we elaborated a taxonomy that lists the 9 most typical design patterns and includes of the formal definition of them using the formal method introduced by Gamma et al. Besides the taxonomy, we provide recommendations on how the specific design patterns can be applied and in what circumstances developers can benefit in Angular application development.

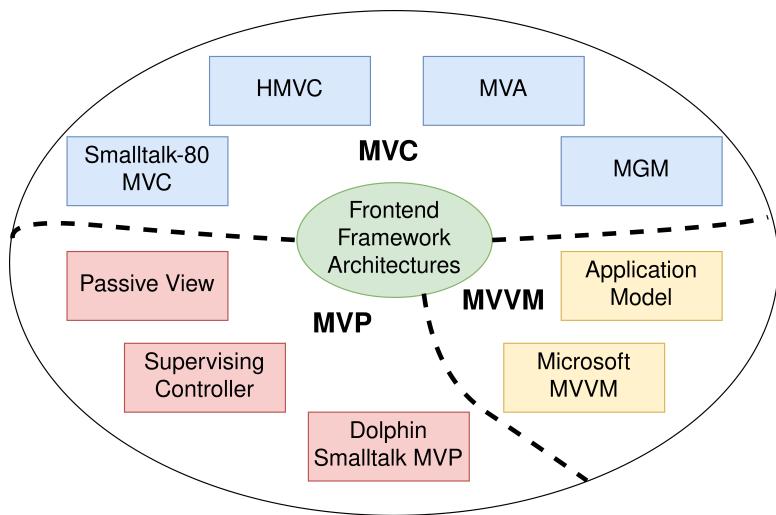


Figure 6: Categories of MVW design patterns and their variants

4.3 Thesis III/3: Detection of Architectural Design Patterns with Large Language Models

I designed a novel prompt engineering methodology for identifying design patterns in Angular projects. I defined a ruleset that helps identifying MVW design patterns in Angular projects. I evaluated both GPT-3.5 and GPT-4 models in rule-based design pattern identification.

Publications related to this thesis: [J5]

Beyond the significance of design patterns, their identification and detection in projects are also important to better understand the source code and see the cohesions among components. Furthermore, valuable statistics can be created based on the evaluation of a number of projects. To develop a neural network for multi-class classification with high accuracy, we would require thousands of projects. The manual labeling is time-consuming and error-prone but automated solutions may help. Today, prompt engineering contains

unrevealed opportunities for extending the use of generative artificial intelligence (AI). Based on our experience gained in architectural design patterns, we constructed a list of rules that help clearly distinguish the MVW design pattern variants in Angular projects. GPT-3.5 and GPT-4 are powerful enough to not only understand our requests but use their knowledge when answering. We presented a ruleset including 30 rules that are independent from each other and help clearly identifying the MVW design pattern variants. We found that GPT-4 performs much better and we were able to achieve an accuracy of 90% in a multi-label classification task. Based on the results, we draw a conclusion of how the various architectural design patterns are applied in practice and gave implementation suggestions in Angular.

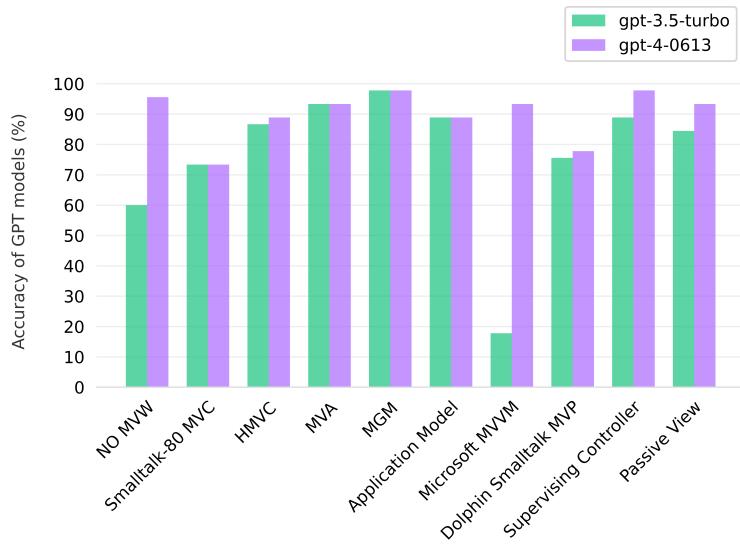


Figure 7: Detection precision of GPT models for MVW design patterns

5 Contributions of the thesis

In the **first thesis group**, I presented different techniques for analyzing data quality and consistency in distributed telemedicine systems efficiently and proposed a taxonomy for managing the trade-off problem in such systems. Detailed discussion can be found in Chapter 3.

I / 1. I explored the complexity of data paths in telemedicine systems and pointed out that data quality and consistency may be reduced due to latency. I proposed using the k -staleness parameter for configuring the cache of distributed telemedicine systems. I gave a formal specification of such a system and elaborated a taxonomy for managing the trade-off problem present in distributed telemedicine systems taking into account the CAP and PACELC Theorems. I made simulations on the model and evaluated them along different k -staleness and latency values.

- I / 2. I presented that modeling and simulation can be an effective combination of analyzing telemedicine systems taking into account the data quality. I showed that a graph representation can be an efficient procedure to evaluate distributed systems as a DAG and significantly reduce the complexity and the size of the data quality related state space.
- I / 3. I proposed approximation techniques to find the best fitting PDFs that can describe the processes of a complex system. I designed a framework that can examine a focused part of the data quality related state space of a system under given circumstances. I demonstrated the effectiveness of the framework with executions of real algorithms. I showed that a significant data quality loss can be observed due to latency in real-time telemedicine systems.

In the **second thesis group**, my contributions are related to the developed SDK, the introduction of WebDAO and the measurements made on development productivity. Chapter 4 provides a comprehensive discussion.

- II / 1. I presented that the front-end and back-end communication can be standardized with the WebDAO design pattern in modern Web development. Especially in telemedicine, the FHIR standard makes a perfect combination with WebDAO. I created usage statistics about FHIR in Web development. I showed that the layered architecture changed due to the aspiration of serverless development and demonstrated how Angular application development utilizes the new architecture.
- II / 2. I developed a SDK for telemedicine application development that supports hybrid cloud approach and applies the WebDAO design pattern. I proposed a method that shows the quantitative extent of development productivity improvement utilizing WebDAO. I made evaluations on the productivity involving real developers and showed that WebDAO can significantly improve development productivity.
- II / 3. I showed a productivity measurement method using repository metrics. Based on the results, I presented that if WebDAO is applied in Angular application development, it can reduce the source code changes and the solutions for the issues require less time.

My contributions to the **third thesis group** are related to NLP methods used in design pattern detection and the achieved results. Chapter 5 incorporates an extensive overview.

- III / 1. I developed and compared two NLP methods for classifying Angular projects and services based on the presence of the WebDAO design pattern. I showed that WebDAO is an implicitly used design pattern in Angular application development and presented that there is a number of projects that do not use services at all.
- III / 2. I introduced the available variants of Model-View-based architectural design patterns in modern Web development. I proposed a taxonomy including formal definitions and recommendations when a specific MVW design pattern is appropriate. I gave Angular use-cases and examples for applying specific design patterns.

III / 3. I designed a novel prompt engineering method for identifying design patterns in Angular projects. I defined a rule set that helps identifying MVW design patterns in Angular projects. I evaluated both GPT-3.5 and GPT-4 models in rule-based design pattern identification.

Table 3 summarizes the relation between the thesis points and the corresponding publications.

Table 3: Correspondence between the thesis points and my publications.

Publication	IF	SJR	Score	Thesis point								
				I/1	I/2	I/3	II/1	II/2	II/3	III/1	III/2	III/3
[J1]	-	Q3	0.75	•								
[J2]	-	Q3	1		•							
[J3]	2.9	Q2	0.75			•						
[J4]	2.6	Q2	0.75				•			•	•	
[J5]	2.9	Q2	0.75								•	•
[J6]	-	Q4	0.75				•	•				
[C1]	-	-	0.48					•				
[FR1]	-	-	-					•				
[FR2]	-	-	-	•								
[FR3]	-	-	-	•								
[FR4]	-	-	-					•				

The author's publications on the subjects of the thesis

Journal publications

- [J1] **Z.R. Jánki** and V. Bilicki. Taxonomy for The Trade-off Problem in Distributed Telemedicine Systems. *Acta Cybernetica*, 25(2), pp. 285–306, 2021. SJR: Q3 (0.75p)
- [J2] **Z.R. Jánki**. A graph-based data quality analysis in distributed telemedicine systems. *Pollack Periodica*, 17(1), pp. 18–23, 2022. SJR: Q3 (1p)
- [J3] **Z.R. Jánki** and V. Bilicki. A Data Quality Measurement Framework Using Distribution-Based Modeling and Simulation in Real-Time Telemedicine Systems. *Applied Sciences*, 13(13), 7548, 2023. IF: 2.9 SJR: Q2 (0.75p)
- [J4] **Z.R. Jánki** and V. Bilicki. The Impact of the Web Data Access Object (WebDAO) Design Pattern on Productivity. *Computers*, 12(8), 149, 2023. IF: 2.6 SJR: Q2 (0.75p)
- [J5] **Z.R. Jánki** and V. Bilicki. Rule-based Architectural Design Pattern Recognition with GPT models. *Electronics*, 12(15), 3364, 2023. IF: 2.9 SJR: Q2 (0.75p)
- [J6] **Z.R. Jánki** and V. Bilicki. Standardized Telemedicine Software Development Kit with Hybrid Cloud Support. *Acta Cybernetica*, 2023. (accepted) SJR: Q4 (0.75p)

Full papers in conference proceedings

- [C1] **Z.R. Jánki**, Z. Szabó, V. Bilicki and M. Fidrich. Authorization solution for full stack FHIR HAPI access. In *Proceedings of the IEEE 30th Neumann Colloquium (NC)*, pp. 121–124, 2017. (0.48p)

Further related publications

- [FR1] **Z.R. Jánki** and V. Bilicki. Full-stack FHIR-based MBaaS with Server- and Client-side Caching Capable WebDAO. In *The 11th Conference of PhD Students in Computer Science*, pp. 192–196, 2018.

- [FR2] **Z.R. Jánki** and V. Bilicki. Crosslayer Cache for Telemedicine. In *The 12th Conference of PhD Students in Computer Science*, pp. 159–163, 2020.
- [FR3] **Z.R. Jánki** and V. Bilicki. The Trade-off Problem in Distributed Telemedicine Systems: A Case-study. In *The 16th Iványi Miklós International PhD & DLA Symposium*, 2020.
- [FR4] **Z.R. Jánki** and V. Bilicki. Domain Specific Semantic Data Model Integration. In *The 13th Conference of PhD Students in Computer Science*, pp. 196–200, 2022.

Further publications

- [F1] Z. Szabó, V. Bilicki, Á. Berta and **Z.R. Jánki**. Smartphone-based data collection with stunner using crowdsourcing: lessons learnt while cleaning the data. In *Proceedings of International Academy, Research, and Industry Association*, 2017. **0.48p**
- [F2] Z. Szabó, V. Bilicki, Á. Berta and **Z.R. Jánki** Smartphone-based data collection with Stunner, the reality of peer-to-peer connectivity and web real-time communications using crowdsourcing: Lessons learnt while cleaning the data. *International Journal On Advances in Software*, vol. 11, no. 1-2, pp. 120–130, 2018.
- [F3] A. Czégel, B. Lakos, **Z.R. Jánki**, A. Bánhalmi and V. Bilicki Elektrokardiográfiai felvételek digitalizálása és kiértékelése okostelefonnal. *XXXII. Neumann Kollokvium*, 2019.
- [F4] **Z.R. Jánki** and V. Bilicki Adatkonzisztencia és adatminőség elosztott telemedicina rendszerekben. *XXXIII. Neumann Kollokvium*, 2020.
- [F5] V. Bilicki, Z. Szabó and **Z.R. Jánki** Számítógép hálózatok programozóknak. *University of Szeged*, 2020.

Total publication score: 5.71

Other References

- [1] Daniel Abadi. Consistency tradeoffs in modern distributed database system design: Cap is only part of the story. 45(2):37–42, 2012.

- [2] Peter Bailis, Shivaram Venkataraman, Michael J. Franklin, Joseph M. Hellerstein, and Ion Stoica. Probabilistically bounded staleness for practical partial quorums. *Proc. VLDB Endow.*, 5(8):776–787, apr 2012.
- [3] Eric Brewer. Cap twelve years later: How the "rules" have changed. *Computer*, 45(2):23–29, 2012.
- [4] Index - fhir v5.0.0. <http://hl7.org/fhir/>.
- [5] M. Fowler. *Patterns of Enterprise Application Architecture*. A Martin Fowler signature book. Addison-Wesley, 2003.
- [6] E. Gamma, R. Helm, R. Johnson, and J. Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional Computing Series. Pearson Education, 1994.
- [7] Inclouded. <http://inclouded.hu/>.
- [8] Mayuram S Krishnan, Charlie H Kriebel, Sunder Kekre, and Tridas Mukhopadhyay. An empirical analysis of productivity and quality in software products. *Management science*, 46(6):745–759, 2000.
- [9] Leslie Lamport, John Matthews, Mark Tuttle, and Yuan Yu. Specifying and verifying systems with tla+. In *Proceedings of the 10th Workshop on ACM SIGOPS European Workshop*, EW 10, page 45–48, New York, NY, USA, 2002. Association for Computing Machinery.
- [10] Harlan D Mills. Software productivity. 1988.
- [11] Mark Richards. *Software architecture patterns*. O'Reilly Media, 2015.
- [12] Reuven Y Rubinstein and Dirk P Kroese. *Simulation and the Monte Carlo method*. John Wiley & Sons, 2016.
- [13] Stefan Wagner and Melanie Ruhe. A systematic review of productivity factors in software development. *CoRR*, abs/1801.06475, 2018.
- [14] Yuan Yu, Panagiotis Manolios, and Leslie Lamport. Model checking tla+ specifications. In *Advanced Research Working Conference on Correct Hardware Design and Verification Methods*, pages 54–66. Springer, 1999.
- [15] Cheng Zhang and David Budgen. What do we know about the effectiveness of software design patterns? *IEEE Transactions on Software Engineering*, 38(5):1213–1231, 2012.

6 Összefoglalás

Az értekezés a modern web-fejlesztésben alkalmazott technikákat és tervezési mintákat tárja fel, valamint különböző telemedicine használati esetek mentén demonstrálja ezek működését és eredményét.

A munka három téziscoport mentén mutatja be a tudományos növumokat. Az első téziscoportban az elosztott telemedicine rendszerek és azok modellezése, valamint szimulációja kerül bemutatásra, ahol az adat áll a fókuszban. A második téziscoport egy szoftverfejlesztői csomagot (SDK-t) mutat be, illetve annak hatásait a fejlesztői produktivitásra. A harmadik téziscoport pedig a modern web-fejlesztésben alkalmazott tervezési mintákat mutatja be, illetve természetes nyelvfeldolgozási (NLP) módszerek segítségével a minták felismerésére adunk megoldást.

Egy rendszer elvárt működéséhez ismernünk kell a konkrét használati esetet. Formális rendszermodell és szimuláció során megfigyeltük, hogy a használati eset által megengedett késleltetés nagyban meghatározza, hogy a rendszer képességeit illetően milyen kompromisszumokat kell kötnünk. Továbbá a szimulációs eredmények azt mutatták, hogy a késleltetés következtében jelentősen romolhat az adatminőség. Ennek vizsgálatára dolgoztam ki egy gráf-alapú klaszterező megoldást, illetve egy emuláción és közelítő módszereken alapuló fókuszált szimulációs környezetet.

Telemedicine fejlesztésein során sikerült jelentős tapasztalatokra szert tenni a modern web-fejlesztés és kapcsolódó területein. A szoftverfejlesztési tervezési minták használatának jelentősége sosem volt kérdéses, azonban az új technológiák és keretrendszerek módosításokat alkalmazhatnak a jól bevált koncepciókon. A WebDAO tervezési minta bemutatásával és alkalmazásával érezhető javulást értünk el a produktivitásban, így ezt kutatásunk során számszerűsítettük is. A kezdeti sikerekre támaszkodva megvizsgáltuk, hogy a világban milyen minőségen van jelen a minta és hogyan hat a produktivitásra.

Végül megvizsgáltuk, hogy a modern web-fejlesztésben a régóta ismert, jól bevált Modell-Nézet-alapú tervezési minták hogyan alakultak át az újítások hatására, illetve különböző NLP technikákkal meglehetősen ígéretes eredményeket értünk el a minták azonosításában.

Mindhárom terület tartogat még nyitott kérdéseket és számos új kutatási irányt nyit meg előttünk.

Nyilatkozat

Jánki Zoltán Richárd *Telemedicine Engineering: Modeling and Analyzing Data Quality in Telemedicine Systems and the Impact of Novel Design Patterns on Productivity in Modern Web Application Development* című PhD disszertációjában a következő eredményekben Jánki Zoltán Richárd hozzájárulása volt a meghatározó:

I/1. tézispont:

1. Adatkonzisztenza és adatminőség mérési módszerek definiálása. Elosztott telemedicina rendszer formális specifikációja. Szimulációs eredmények az adatkonzisztenzia változására a gyorsítótár k -romlási paraméterének változtatása mentén. Elosztott telemedicina rendszerek taxonómijának megalkotása.
 - Z. R. Jánki and V. Bilicki, *Taxonomy for Trade-off Problem in Distributed Telemedicine Systems*, *Acta Cybernetica*, vol. 25, no. 2, pp. 285–306, Aug. 2021. SJR: Q3
2. Adatkonzisztenzia vizsgálatának formális leírása elosztott telemedicina rendszereknél. Elosztott telemedicina rendszerekhez paraméterezhető gyorsítótárazási modell implementálása. Szimulációs eredmények az adatkonzisztenzia változására a gyorsítótár k -romlási paraméterének változtatása mentén.
 - Z.R. Jánki and V. Bilicki. *Crosslayer Cache for Telemedicine*. In *The 12th Conference of PhD Students in Computer Science*, pp. 159—163, 2020.
3. A CAP- és PACELC-tételben ismertetett kompromisszum problémájának (trade-off problem) demonstrálása egy telemedicina használati esettel.
 - Z. R. Jánki and V. Bilicki, *The Trade-off Problem in Distributed Telemedicine Systems: A Case-study*, In *The 16th Iványi Miklós International PhD & DLA Symposium*, 2020.

I/2. tézispont:

1. Formális telemedicina rendszerespecifikáció készítése. Gráf-alapú adatminőség mérése elosztott telemedicina rendszereknél. DAG struktúra megállapítása és a gráf gyengén összefüggő komponenseinek klaszterezése az adatminőségi értékek mentén.
 - Z. R. Jánki, *A graph-based data quality analysis in distributed telemedicine systems*, *Pollack Periodica*, vol. 17, no. 1, pp. 18–23, 2022. SJR: Q3

I/3. tézispont:

1. A kiválasztott használati eset rendszerének formális leírása. A valós idejű telemedicina rendszereknél potenciálisan felmerülő adatminőség csökkenés problémájának felvázolása. Emulációs környezet kiépítése a rendszer folyamatainak futtatásához. Emulációs eredmények mentén közelítő módszerek alkalmazásával egy fókuszált állapottér kialakítása. Szimulációs keretrendszer megvalósítása az állapottér vizsgálatához és az eredmények kiértékelése.
 - Z. R. Jánki and V. Bilicki, *A Data Quality Measurement Framework Using Distribution-Based Modeling and Simulation in Real-Time Telemedicine Systems*, *Applied Sciences*, vol. 13, no. 13, 7548, 2023. IF: 2.9 SJR: Q2

II/1. tézispont:

1. WebDAO tervezési minta bemutatása a modern Web fejlesztésben.
 - Z. R. Jánki and V. Bilicki, The Impact of the Web Data Access Object (WebDAO) Design Pattern on Productivity, Computers, vol. 12, no. 8, 149, 2023. IF: 2.6 SJR: Q2
2. Statisztikák készítése az FHIR szabvány használatáról nyílt forrású projektekben. Az Inclouded SDK képességeinek bemutatása: hibrid felhő támogatás, polyglot perzisztencia megvalósítása, szabványos FHIR interfések implementálása.
 - Z. R. Jánki and V. Bilicki, Standardized Telemedicine Software Development Kit with Hybrid Cloud Support, Acta Cybernetica, 2023. (elfogadva) SJR: Q4

II/2. tézispont:

1. A WebDAO tervezési minta implementálása az Inclouded SDK formájában. A produktivitás javulásának demonstrációja valós fejlesztők bevonásával és az eredmények kiértékelése.
 - Z. R. Jánki and V. Bilicki, Standardized Telemedicine Software Development Kit with Hybrid Cloud Support, Acta Cybernetica, 2023. (elfogadva) SJR: Q4
2. Mobile Backend as a Service (MBaaS) megvalósítás egy SDK formájában szabványos FHIR interfésszel.
 - Z. R. Jánki, Z. Szabó, V. Bilicki and M. Fidrich, Authorization solution for full stack FHIR HAPI access, 2017 IEEE 30th Neumann Colloquium (NC), Budapest, Hungary, 2017, pp. 121–124, doi: 10.1109/NC.2017.8263266.
3. Gyorsítótárazást támogató MBaaS megvalósítás a WebDAO tervezési minta alkalmazásával.
 - Z. R. Jánki, and V. Bilicki, Full-stack FHIR-based MBaaS with Server- and Client-side Caching Capable WebDAO, The 11th Conference of PhD Students in Computer Science (CSCS), Szeged, Hungary, 2018, pp. 179–183.
4. FHIR szabvány bővíthetőségének támogatása NLP megoldásokkal.
 - Z. R. Jánki and V. Bilicki, Domain Specific Semantic Data Model Integration, In The 13th Conference of PhD Students in Computer Science, 2022. pp. 196–200.

II/3. tézispont:

1. A WebDAO tervezési minta alkalmazása a modern web-fejlesztésben. A WebDAO produktivitásra gyakorolt hatásaival kapcsolatos eredmények az Angular alkalmazásfejlesztésben.
 - Z. R. Jánki and V. Bilicki, The Impact of the Web Data Access Object (WebDAO) Design Pattern on Productivity, Computers, vol. 12, no. 8, 149, 2023. IF: 2.6 SJR: Q2

III/1. tézispont:

1. A WebDAO tervezési minta osztályozására implementált Random Forest osztályozó. A WebDAO tervezési minta osztályozására implementált BERT modell. A WebDAO Angular keretrendszerben történő alkalmazásával kapcsolatos statisztikai eredmények.

- Z. R. Jánki and V. Bilicki, The Impact of the Web Data Access Object (WebDAO) Design Pattern on Productivity, Computers, vol. 12, no. 8, 149, 2023. IF: 2.6 SJR: Q2

III/2. tézispont:

1. Model-View-Whatever (MVW) tervezési mintacsalád taxonómiájának megalkotása, a minták formális definiálása. MVW tervezési minták használatára javaslatok és implementációs minták kidolgozása Angular keretrendszerben.

- Z. R. Jánki and V. Bilicki, Rule-based Architectural Design Pattern Recognition with GPT models, Electronics, vol. 12, no. 15, 3364, 2023. IF: 2.9 SJR: Q2

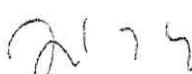
III/3. tézispont:

1. Angular keretrendszerre épített szabályrendszer kidolgozása, amely az MVW tervezési mintákat egyértelműen definiálja. GPT-modellek kiértékelése és összehasonlítása az MVW tervezési minták azonosítására Angular projektekben.

- Z. R. Jánki and V. Bilicki, Rule-based Architectural Design Pattern Recognition with GPT models, Electronics, vol. 12, no. 15, 3364, 2023. IF: 2.9 SJR: Q2

Ezek az eredmények Jánki Zoltán Richárd PhD disszertációján kívül más tudományos fokozat megszerzésére nem használhatók fel.

Szeged, 2023. augusztus 8.



Jánki Zoltán Richárd
jelölt

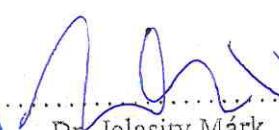


Dr. Bilicki Vilmos
témavezető

Az Informatika Doktori Iskola vezetője kijelenti, hogy jelen nyilatkozatot minden társzerző-höz eljuttatta, és azzal szemben egyetlen társzerző sem emelt kifogást.

Szeged, 2023. 08. 21.





Dr. Jelasity Márk
SZTE Informatika Doktori Iskola vezetője

